

Metabarcoding soil microarthropods for soil quality assessment: Importance of integrated taxonomy, phylogenetic marker selection and sampling design

by

Jesse Frank James Hoage

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science (MSc) in Biology

The Faculty of Graduate Studies
Laurentian University
Sudbury, Ontario, Canada

© Jesse Frank James Hoage, 2018

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian University/Université Laurentienne
Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis Titre de la thèse	Metabarcoding soil microarthropods for soil quality assessment: Importance of integrated taxonomy, phylogenetic marker selection and sampling design	
Name of Candidate Nom du candidat	Hoage, Jesse	
Degree Diplôme	Master of Science	
Department/Program Département/Programme	Biology	Date of Defence Date de la soutenance September 10, 2018

APPROVED/APPROUVÉ

Thesis Examiners/Examineurs de thèse:

Dr. Nathan Basiliko
(Co-Supervisor/Co-directeur de thèse)

Dr. Lisa Venier
(Co-Supervisor/Co-directrice de thèse)

Dr. Nadia Mykytczuk
(Committee member/Membre du comité)

Dr. Yves Alarie
(Committee member/Membre du comité)

Dr. Pedro Antunes
(External Examiner/Examineur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. David Lesbarrères
Monsieur David Lesbarrères
Dean, Faculty of Graduate Studies
Doyen, Faculté des études supérieures

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

I, **Jesse Hoage**, hereby grant to Laurentian University and/or its agents the non-exclusive license to archive and make accessible my thesis, dissertation, or project report in whole or in part in all forms of media, now or for the duration of my copyright ownership. I retain all other ownership rights to the copyright of the thesis, dissertation or project report. I also reserve the right to use in future works (such as articles or books) all or part of this thesis, dissertation, or project report. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that this copy is being made available in this form by the authority of the copyright owner solely for the purpose of private study and research and may not be copied or reproduced except as permitted by the copyright laws without written authority from the copyright owner.

Abstract

Soil microarthropods are ubiquitous and ecologically significant in terrestrial environments. Collembola and Oribatida are the two most abundant and diverse representatives of microarthropods and are commonly targeted biological indicators of soil quality. Traditional methods for studying these groups have provided taxonomic and functional data for individuals but are inefficient for large scale biomonitoring applications. The adoption of molecular methods like metabarcoding is predicted to improve the efficiency of biomonitoring microarthropods but can be limited by the availability of reference specimens in sequence databases, and the importance of barcode selection and sampling design for this approach in plot-level experiments has not been explored. Demonstrated here, the inclusion of locally derived specimen barcodes to reference libraries significantly improved the quality of metabarcoding data. Also, targeting two barcodes (18S and COI) improved microarthropod richness estimates, and 10 samples with >10 m separation between each is recommend to increase the proportion of diversity detected.

Keywords

Microarthropods, Collembola, Oribatida, metabarcoding, DNA barcoding, richness estimate, spatial autocorrelation, integrated taxonomy

Co-Authorship Statement

Drs. Nathan Basiliko and Lisa Venier made large contributions to the thought behind this thesis and thoroughly edited the entire work. Dr. Teresita Porter oversaw the high-throughput sequencing, conducted the bioinformatics processing of the raw sequence data, authored the bioinformatic workflow and contributed edits throughout the work. Dr. Laurent Rousseau contributed Collembola specimens that he identified from his thesis work for the local reference library.

Acknowledgments

First, I would like to acknowledge Drs. Nathan Basiliko and Lisa Venier, for being patient, supportive, and committed supervisors. They have allowed me to learn from my mistakes, encouraged me to try new things, and are perfect examples of professionals and mentors. I also would like to thank my committee members Drs. Nadia Mykytczuk and Yves Alarie for their patience as well and supportive contributions. Three other who must be acknowledged are Kerrie Wainio-Keizer, Terri Porter and Laurent Rousseau. Kerrie was my mentor through much of my work in the field and always made me feel comfortable during our work-ations at Island Lake. Terri has been extremely generous and supportive throughout my thesis, and the quality of this work would not have been the same without the effort from her. Laurent welcomed me as the first graduate student I worked with while I was in Sault. Ste. Marie where he introduced me to these organisms and has always been supportive whenever we have had the chance to connect. Also, to the entire crew at the Living with Lakes Centre, it has been an amazing experience working in this atmosphere with everyone who has come and gone. One more individual I would like to acknowledge is my high-school biology teacher Mr. Aleksander Stosich. It was during his classes where my interest in science was found, and this was encouraged by his enthusiasm for the field and commitment to students. Finally, my family, friends, Megan, and all my cats have been the greatest motivation and support group I could ask for. Without the influence of all these people, I would not have achieved this goal that I set many years ago.

May the force be with you all.

Table of Contents

Abstract.....	iii
Co-Authorship Statement.....	iv
Acknowledgments.....	v
Table of Contents.....	vi
List of Tables	viii
List of Figures.....	ix
Introduction.....	1
1.1 Ecological Significance of Soil Microarthropods.....	2
1.2 Microarthropods as Biological Indicators of Soil Quality.....	5
1.3 Issues with Traditional Methods for Microarthropod Biomonitoring	7
1.4 Molecular Methods for Biodiversity Assessment.....	8
1.5 Thesis Objectives	11
2 Methods.....	13
2.1 Study Site	13
2.2 Collection and Identification of Collembola Specimens	14
2.3 DNA Extraction from Collembola Specimens	14
2.4 PCR of COI Barcodes from Collembola Specimen DNA	15
2.5 Sanger Sequencing and Proofreading of Collembola COI Barcodes	16
2.6 Identification of Collembola COI Barcodes with the Barcode of Life Database ..	16
2.7 Sampling and Processing of Soil Samples.....	17
2.8 eDNA Extraction from Soil Samples.....	18
2.9 PCR Amplification of DNA Barcodes from Soil eDNA	19
2.10Illumina MiSeq Amplicon Sequencing DNA Barcodes	20

2.11 Bioinformatics and Data Analysis	21
3 Results	24
3.1 Morphological and COI Barcode Identification of Local Collembola Specimens	24
3.2 Impact of Local Collembola COI Barcodes on OTU Classification Confidence .	25
3.3 Metabarcoding Sequencing Summary, OTU Distributions and Barcode Specificity.....	25
3.4 Identification-Sensitivity of Metabarcoding Compared to Traditional Methods .	27
3.5 Sampling Depth and Spatial Dependency.....	28
4 Discussion	29
4.1 Barcoding Local Specimens Can Improve Global Taxonomy and Ecology	31
4.2 Phylogenetic Barcode Performance Varies Across Taxa	36
4.3 Number of Samples and Spatial Arrangement Can Impact Metabarcoding Results	40
5 Conclusions	42
References	55

List of Tables

Table 1: Morphological species and top match in BOLD based on COI barcode identification of local Collembola specimens. Representation of morphological species in BOLD is given as +/-. BINs are provided for COI barcodes with $\geq 95\%$ similarity to top match.	44
Table 2: Sequencing summary of the 4 DNA barcodes. Values are represented as all 44 samples cumulatively (total) and the average per sample. Read numbers are for paired and quality filtered results.	45
Table 3: Comparison of unique genera identified by each barcode for the three targeted taxonomic groups. Collembola and Oribatida are represented by the barcodes CO1-F230, CO1–BR5, and 18S rRNA. Fungi are represented by the 18S rRNA and ITS2 barcodes.....	45
Table 4: Comparison of Collembola and Oribatida genera identified from Island Lake with a traditional approach from Rousseau et al. (2018) and through metabarcoding. We used the minimum COI bootstrap proportion (BP) cutoffs suggested by Porter and Hajibabaei (2018b). Bracketed values indicate genera shared with morphological data set. Values in parentheses indicate number of taxa unique to metabarcoding data sets.	46

List of Figures

Figure 1: Schematic diagram for spatial sampling of exploratory plot at Island Lake site. All distances are in metres.	46
Figure 2: Comparison of bootstrap support values for taxonomic identification of Collembola reads between the unmodified reference library and one supplemented with the local Collembola COI barcodes. The vertical line indicates the recommended bootstrap support cutoff (30%) for 99% confident sequence identification.	47
Figure 3: Distribution of the relative number of OTUs classified for each barcode, separated by taxonomic kingdom. ITS2 was not included as all OTUs belonged to Fungi.	48
Figure 4: Rarefaction curve with 95% confidence limits for the accumulation of total OTU richness with increasing number of samples for each barcode. Vertical line indicates the average number of samples at which $\geq 90\%$ of species richness is achieved.	49
Figure 5: Rarefaction curve with 95% confidence limits for the accumulation of total OTU richness with increasing number of samples for Arthropoda and Fungi OTUs. Vertical lines indicates the number of samples at which $\geq 90\%$ of species richness is achieved.	50
Figure 6: Rarefaction curve with 95% confidence limits for the accumulation of total OTU richness with increasing number of samples for Collembola and Oribatida OTUs. Vertical lines indicates the number of samples at which $\geq 90\%$ of species richness is achieved.	51
Figure 7: Histogram of all spatial distances between pairwise samples from exploratory plot, summarized into 2 m intervals.	52
Figure 8: Relationship between Jaccard distance given spatial distance for each taxonomic rank with 95% confidence limits. Curve calculated with generalized additive model.	53
Figure 9: Correlogram of Mantel r values of 20 distance classes for each taxonomic rank. Mantel r values calculated from community dissimilarity and spatial distances. Solid points indicate	

significant p-values (≤ 0.05) and the size of points indicate relative number of values within each distance class.....	54
--	----

Introduction

Soil biota are an essential component of all terrestrial ecosystems, yet are considered one of the remaining frontiers in biodiversity research (Orgiazzi et al., 2015). This habitat hosts organisms from all three domains of life (Eukarya, Bacteria, Archaea), including representatives from every eukaryotic kingdom (Bardgett and van der Putten, 2014). Biodiversity estimates in soil range from 10^4 to 10^7 bacterial species in a 1 cm^3 soil sample and up to thousands of invertebrate species within a 1 m^2 plot (Decaëns, 2008; Fierer et al., 2009, 2007; Roesch et al., 2007). Various taxonomic groups have been utilized as biological indicators for assessing soil quality in forestry. In addition to abiotic conditions of soil, the flora and fauna are ideal indicators of soil quality as they directly influence the chemistry and structure of their environment, have relatively rapid turnover rates, and are particularly sensitive to environmental disturbances (Karimi et al., 2017; Socarrás, 2013). Along with inconsistencies among taxonomists and soil ecologists, many factors inherent to the nature of the organisms and their habitat complicate the traditional collection and identification of samples (André, 2002; Rougerie et al., 2009). Correcting this impediment requires the integration of traditional methods and modern technologies in a coordinated multidisciplinary approach to the study of soil ecology (Cristescu, 2014). Novel molecular based techniques (e.g. DNA barcoding/metabarcoding, metagenomics, transcriptomics) have been developed to improve our capability of investigating microbial communities. In the context of soil microarthropods and other fauna, these technologies have a similar potential, as they can provide an alternative to the impediments and biases of isolation and identification in this field.

Soil microarthropods (e.g. Collembola and Oribatida) represent an ideal niche in biodiversity and ecology research for the application of a coordinated approach using morphological and molecular data in tandem. These organisms fall on the boundary of being large enough for specimens to be physically sorted and described in detail with the use of a microscope, yet small enough for their diversity to be accurately detected with modest samples of soil or litter, amenable to environmental DNA (eDNA) extraction (Taberlet et al., 2012a). With an integrated approach, an extensive and detailed molecular reference library can be established by traditional and molecular taxonomists to provide ecological context for the data generated from molecular based biodiversity assessments. Here, I propose a protocol for soil biodiversity assessment that enhances DNA metabarcoding by supplementing global sequence reference libraries with morphologically and genetically characterized specimens that have been locally collected.

1.1 Ecological Significance of Soil Microarthropods

Collembola and Oribatida are the two most abundant taxonomic groups of soil microarthropods and are considered the model representatives for soil mesofauna (organisms ~0.1 – 2 mm in size) due to their ubiquity, diversity, and ecological significance (Behan-Pelletier, 1999; Rusek, 1998). They are globally present in terrestrial ecosystems, from polar to equatorial regions, and can be found throughout the soil profile, in above-ground litter and woody-debris, on vegetation, bark and in tree canopies. Abundance and diversity in these groups can have a wide range depending on the environment, with up to 10^4 individuals representing 10 – <100 species per m^2 (Bardgett and van der Putten, 2014). Morphologically, Collembola are a class of soft-bodied hexapods that can possess a ventral, forked jumping appendage known as a

furca, from which their common name (springtails) is derived, while Oribatida are an acarine order of heavily sclerotized mites (Hopkin, 1997; Walter and Proctor, 2013). Both groups can reproduce sexually or asexually (via parthenogenesis) but Oribatida are generally considered more as K-selected organisms than Collembola, with lower metabolic and reproductive rates (Gulvik, 2007). Because of this, Collembola are typically able to colonize soils earlier than Oribatida (Ingimarsdóttir et al., 2012; Ojala and Huhta, 2001).

The taxonomic diversity of Collembola and Oribatida is also accompanied by high trophic diversity. They are commonly grouped together simply as fungivores and decomposers, however food web studies based on morphological characteristics and stable isotope fractionation consistently propose a range of 4-5 feeding guilds that generally include predators, fungivores, herbivores, and decomposers (Malcicka et al., 2017; Perdomo et al., 2012; Potapov et al., 2016). A high level of general grazing is also commonly suggested, predominantly between herbivores, fungivores and decomposers (Oelbermann and Scheu, 2010). This is often attributed to the presence of fungi on detritus and in the rhizosphere, resulting in the deliberate or coincidental consumption of both substrates (Harrop-Archibald et al., 2016; Maraun et al., 2003a). Within fungivores, feeding experiments and stable isotope analyses show that some species selectively feed on certain fungal groups or taxa whereas others show no preference (Potapov and Tiunov, 2016; Schneider et al., 2004; Staaden et al., 2011). Selective feeding towards dominant fungi promotes diversity in the fungal community and can limit the proliferation of pathogenic groups (Crowther et al., 2013; Friberg et al., 2005)

The activities of microarthropods have multiple important dynamic effects on the abiotic components of soil and on other biotic communities that directly and indirectly alter quality and fertility. Shredding of litter by primary decomposing microarthropods has broad consequences

on its decomposition. Comminution of litter increases the surface area available for microbial colonization (Bradford et al., 2002; Soong et al., 2016) and digestion of particles alters the litter chemistry and inoculates fecal pellets with microbes (Buse et al., 2014; Wickings and Grandy, 2011), enhancing microbial growth and nutrient cycling (Heneghan and Bolger, 1998).

Fungivorous microarthropods have large impacts on soil dynamics as grazing liberates nutrients from senescent fungal hyphae and stimulates compensatory fungal growth and metabolism, increasing fungal biomass and further enhancing microbial abundance and productivity (A'Bear et al., 2014; Cole et al., 2003; Soong et al., 2016). Additionally, microarthropods distribute microbial cells and fungal spores throughout the soil as they migrate (Lilleskov and Bruns, 2005; Renker et al., 2005). These activities help develop the physical structure of soil through aggregation of fecal pellets and litter particles and stabilization by fungal hyphae, promoting soil aeration and hydration (Maaß et al., 2015; Siddiky et al., 2012).

With all these resulting changes to soil, above and belowground plant productivity has been shown to be influenced by soil microarthropod abundance and diversity. Higher total biomass or abundance of microarthropods typically corresponds to enhanced net primary productivity, with a particularly strong influence in conifer dominated ecosystems (Cole et al., 2003; Sackett et al., 2010). Belowground, root biomass and colonization by arbuscular mycorrhizal fungi increase with microarthropod abundance (Chen et al., 2017; Hishi and Takeda, 2008; Steinaker and Wilson, 2008). Coinciding with this, higher foliar nutrients and an increase in foliar herbivory is also positively correlated with soil microarthropod abundance and diversity (Callejas-Chavero et al., 2015; Soler et al., 2012). Finally, abundance of microarthropods can enhance succession of grassland plant communities by feeding on the roots of dominant early

stage plants, indirectly providing advantage to late stage successional plants (De Deyn et al., 2003).

1.2 Microarthropods as Biological Indicators of Soil Quality

Microarthropod communities have been shown to be useful indicators of soil quality under various land use practices due to their role in enhancing soil functional processes and sensitivity to perturbation. Their abundance and diversity allow small and large changes in the communities to be observed and interpreted as changes in the functional nature of the soil system. This has been taken advantage of in many instances to measure the health/succession of an environment/ecosystem or the significance of natural and artificial impacts and as an indicator of the health of connected communities (lower/higher trophic levels i.e. flora/larger fauna). A variety of biological indexes based on microarthropod populations have also been proposed and implemented to quantify soil quality (e.g. Caruso et al., 2007; Parisi et al., 2005; Yan et al., 2012).

Forestry is an economically and environmentally important industry in Canada and globally that relies on maintaining healthy soils in the face of intensified harvesting and climate change (Achat et al., 2015; Kirilenko and Sedjo, 2007; Thiffault et al., 2010). In many studies microarthropods have been successfully used to measure the level of impact that various harvesting and site preparation practices can have on soil quality and the rate of succession after these disturbances. In experimental gradients of typical harvesting, impacts such as biomass removal, soil compaction and mechanical perturbation there is generally an overall reduction in microarthropod abundance and diversity as the intensity of disturbance increases (Battigelli et

al., 2004; Maraun et al., 2003b; Rousseau et al., 2018). These treatments produced environments with reduced soil moisture content from the loss of ground cover, reduced soil pore space, and decreased fungal biomass which resulted in a shift towards generalist and parthenogenic species, with selection against sexual reproductive, large epiedaphic, some euedaphic, and fungivorous species. In forestry, there are different harvesting and site preparation practices that vary in the intensity of these impacts and microarthropod communities are commonly targeted to help measure the sustainability of these practices. Different responses of the microarthropod community have been shown between partial- and clear-cutting methods (Addison and Barber, 1997; Lindo and Visser, 2004; Siira-Pietikäinen et al., 2001), stem-only and whole-tree harvesting (Battigelli et al., 2004; Rousseau et al., 2018), and different site preparation techniques including prescribed burning (Berch et al., 2007; Bird et al., 2004; Malmström et al., 2009). As well, responses of the belowground community have been measured many years after disturbance, even when aboveground productivity and other abiotic factors are similar (Addison et al., 2003; Chauvat et al., 2007; Farská et al., 2014; Malmström, 2012; Siira-Pietikäinen and Haimi, 2009).

Microarthropod communities have also been successfully implemented as biological indicators of soil quality in other contexts. Similar to forestry, agriculture is another globally significant industry where maintaining soil quality is essential, and microarthropods have been used to identify appropriate harvesting and crop management strategies that sustain soil productivity (Bedano et al., 2016; Coudrain et al., 2016; Crotty et al., 2015; D'Annibale et al., 2017). Species of Oribatida and Collembola (*Oppia nittens* and *Folsomia candida*/*F. fimetaria* respectively) are utilized as standard toxicological indicators of soil contamination (Krogh and Miljøundersøgelser, 2009; Princz et al., 2010). Variations in land-management, and geological

history of sites with changes dating from 100s-1000s years old has also been differentiated using microarthropod communities (Birkhofer et al., 2016; Bokhorst et al., 2017; de la Peña et al., 2016; Zaitsev et al., 2013).

1.3 Issues with Traditional Methods for Microarthropod Biomonitoring

Traditional morphological classification and enumeration of soil microarthropods requires the isolation of individuals from the soil matrix. These collection methods vary widely in success and sensitivity, often requiring different methods to target different taxonomic groups. Collection of microarthropods in the field using pitfall traps and litterbags is common, but there is large variability in the effectiveness across taxa and consistency between samples (Baini et al., 2016; Prasifka et al., 2007). The collection of bulk soil and litter is another common sampling method where microarthropods are later extracted from the bulk samples in a controlled setting. There are a variety of extraction techniques, most reliant on establishing a temperature and/or humidity gradient (e.g. Berlese-Tullgren and McFayden funnels) or flotation (e.g. with sugar/salt solutions or heptane flotation) using sometimes expensive equipment (McSorley and Walter, 1991; Yi et al., 2012). Again, there are issues with all of these methods: there is a large variability in the extraction efficiency across taxa, the time between field sampling and extraction that can result in the mortality of individuals prior to collection, and there is a lack of consistency between studies regarding depth and volume of samples used (André, 2002). Because of this variability, the resulting communities can largely depend on both the target groups and on experimenters' available equipment and preference.

After the organisms are isolated, the process of identification is both time and resource intensive. For soil microarthropods, taxonomists capable of identifying specimens accurately to a genus or species level are rare, and can take significant amounts of time to identify inherently difficult specimens (deWaard et al., 2009; Ji et al., 2013; Porco et al., 2013). With the possibility of $>10^4$ extracted individuals per m^2 of soil (Maraun et al., 2003b), the time required to generate biodiversity estimates for even small-scale studies is an impediment. Recent studies of small-scale (plots $<100 \text{ m}^2$) spatial patterns of Collembola communities required 20-35 samples (1000 cm^3 per sample) to capture $>90\%$ of the total species richness and found that the communities were positively autocorrelated at scales of 0.5-1.5 m, and show a tendency to aggregate across a plot, with significant levels of heterogeneity between samples (Dirilgen et al., 2018; Widenfalk et al., 2016). Other factors impacting the efficiency of traditional methods relying on morphological identification is the lack of distinguishing features in immature life cycle stages and the issue of cryptic morphology between closely related taxa. When molecular techniques are applied, it is commonly discovered that specimens long believed to be representatives of the same species can differ significantly at the phylogenetic level (Hebert et al., 2004; Porco et al., 2012; Zhang et al., 2015). This can lead to differences in biodiversity estimates and limit the effectiveness for using microarthropods as biological indicators in large-scale applied studies.

1.4 Molecular Methods for Biodiversity Assessment

Metabarcoding is a term that has come to refer to identification of taxa using short phylogenetic marker gene sequences (DNA barcode) at the community level (Hebert et al., 2003a; Taberlet et al., 2012b). Metabarcoding can be used with eDNA extracted directly from a bulk environmental sample such as soil and, in principle, can rapidly generate large quantities of

digital biodiversity data through high-throughput sequencing (HTS) that can identify individual taxa in a community, assess phylogenetic relationships between individuals and groups, and compare communities across samples or sites (Ficetola et al., 2008; Gibson et al., 2015; Ji et al., 2013; Taberlet et al., 2012b; Yu et al., 2012). At the individual-specimen level, DNA barcodes have been used to identify and discriminate cryptic species and specimens at immature growth stages (Hebert et al., 2004; Hogg and Hebert, 2004; Maraun et al., 2004; Orgiazzi et al., 2015; Rougerie et al., 2009). The cytochrome c oxidase subunit 1 (COI) mitochondrial DNA (mtDNA) gene has been largely utilized as the barcode for metazoan phylogenetic identification (Hebert et al., 2003b; Ratnasingham and Hebert, 2007), although other nuclear DNA markers are used as well; the 18S and 28S ribosomal RNA (rRNA) genes are broadly used for eukaryotes, and the internal transcribed spacer (ITS) regions of the rRNA cistron are used for some groups, although the ITS regions are specifically considered the phylogenetic barcode for fungi (Geisen et al., 2015; Hamilton et al., 2009; Schoch et al., 2012). COI has been favoured over the rRNA barcodes due to its clonal inheritance and higher rate of mutation, which limits the possibility of recombination events and provides levels of interspecific sequence divergence that can facilitate discrimination between species (Birky, 2001; Hebert et al., 2003b; Knowlton and Weigt, 1998; Tang et al., 2012). Issues regarding the use of COI have been raised, mainly the prevalence of “COI-like” artefacts in reference databases and the lack of true “universal” primer binding sites (Deagle et al., 2014; Galtier et al., 2009), however a comprehensive research plan with appropriate quality controls and selection of primer sets these concerns can be overcome (Cannon et al., 2016; Drummond et al., 2015; Hajibabaei et al., 2005).

There are many technical issues with using a metabarcoding approach at present. The amplification of target sequences through PCR is an important step and the products can be

influenced by multiple factors. Primer selection and binding efficiency may be the largest influencing factor, as optimal barcode selection can vary across taxa and sequences from undiscovered taxa can not be incorporated into primer design, biasing primer binding efficiency towards known taxa (Anslan and Tedersoo, 2015; Cannon et al., 2016; Derycke et al., 2010; Drummond et al., 2015; Lehmitz and Decker, 2017; Op De Beeck et al., 2014; Schoch et al., 2012). Further bias in the PCR step results from reaction stochasticity and differences in template base composition that can alter the relative abundances of taxa due to uneven amplification (Aird et al., 2011; Ficetola et al., 2015; Polz and Cavanaugh, 1998; Schmidt et al., 2013; Suzuki and Giovannoni, 1996). Outside of PCR, other biases can be attributed to eDNA extraction efficiency due to abiotic factors and protocol selection (Donn et al., 2008; Krsek and Wellington, 1999; McKee et al., 2015; Miller et al., 1999; Roh et al., 2006; Sagova-Mareckova et al., 2008; Técher et al., 2010; Terrat et al., 2015), and biomass variability between different body sizes and abundances that result in disproportionate contributions to the eDNA template (Elbrecht et al., 2017; Elbrecht and Leese, 2015). Challenges in bioinformatic processing and the constant development of new technologies and methods also reduce the consistency of results across studies (Coissac et al., 2012; Ficetola et al., 2015; Gibson et al., 2014; Hajibabaei et al., 2005; Yang et al., 2013; Yang and Rannala, 2016). Finally, because taxa are often underrepresented in barcode reference libraries, a large proportion of data generated from metabarcoding lack species level or higher identification, and instead are classified as operational taxonomic units (OTUs), exact sequence variants (ESVs), barcode index numbers (BINs), or species hypotheses (SHs) depending on the taxonomic group, barcode marker, or sequence clustering method used (Blaxter et al., 2005; Callahan et al., 2017; Curry et al., 2018; Kõljag et al., 2013; Ratnasingham and Hebert, 2013). Further developments in -omics fields

might resolve some technical biases in the near future (Simon and Daniel, 2011; Swenson and Jones, 2017; Tedersoo et al., 2018, 2015; Zhou et al., 2013). Integrating traditional taxonomy with the expanding barcoding efforts will increase the depth and quality of biodiversity reference libraries (Cristescu, 2014; Jinbo et al., 2011; Rougerie et al., 2009). For metabarcoding, using multiple gene targets simultaneously may help capture more taxa/microarthropod biodiversity in a single sample (Gibson et al., 2014), however the impact that this would have in an in-situ biodiversity monitoring or environmental impact study is not well understood. As previously described, traditional methods have shown that soil microarthropod communities have high levels of heterogeneity and limited spatial autocorrelation across small-scale plots. It is unclear whether a molecular approach will be similarly affected by this variability and if sampling designs targeting microarthropods should be modified based on the method used for biodiversity estimation.

1.5 Thesis Objectives

Given the challenges noted above with classical morphological approaches and newer DNA barcoding and metabarcoding approaches, research that combines both approaches in a tandem method for soil microarthropod community studies is warranted. Comparisons between the approaches have been performed in a variety of contexts with mixed results (Deiner et al., 2017; Emilson et al., 2017). Similar tandem approaches have been used in conservation science for both invasive species and parasite detection (Barnes and Turner, 2016; Comtet et al., 2015; Smith and Fisher, 2009) and have been used for taxonomic clarification and revisions of individual cryptic taxa (Porco et al., 2010). In the context of biodiversity characterization or biomonitoring at larger scales (e.g. relevant to the scale of boreal forest disturbance in Canada)

the time and expertise requirements make classical approaches unreasonable while metabarcoding is hindered by the underrepresentation of diversity in sequence reference libraries, making links to species and functions difficult. Although reference databases are constantly expanding (Porter and Hajibabaei, 2018a), it is likely that specimens missing from these databases are still common. Therefore, combining the two approaches can result in a greater understanding than either can achieve on its own.

The objective of this master's thesis is to develop a new strategy to survey soil microarthropod communities for soil quality assessments that utilizes both traditional and modern methods in a coordinated approach. This proposed strategy is based on three main questions. First, can public reference libraries be amended with sequences from morphologically classified local specimens to significantly improve the identification confidence of OTUs derived from metabarcoding data? This question has not yet been addressed for terrestrial arthropods in the scholarly literature, and here I conduct the first effort to do so. Because soil fauna are poorly represented in reference libraries, the addition of any novel sequences from local specimens with verified taxonomy to type specimens from reference libraries is expected to have a positive effect on taxa classification in the metabarcoding data where these local sequences are likely to be found. Applying this approach will also help to increase the diversity of microarthropods in reference libraries (Coward *et al.*, 2015). Second, using three DNA barcodes (COI, 18S rRNA, ITS2), can both the microarthropod and fungal communities be more accurately estimated from standard soil eDNA samples, and how consistent are these microarthropod estimates with results from traditional methods? As fungi are an important food source for microarthropods and also have a large influence on decomposition and soil quality, the ability to assess both communities together would be beneficial for biomonitoring. Because DNA barcode selection introduces a

significant sensitivity bias across taxa, utilizing multiple barcodes is an approach that can potentially compensate for this. As molecular and traditional methods are often inconsistent, it is difficult to determine which estimate is more accurate. With multiple barcodes some of these inconsistencies could be decreased. Third, how does the number and arrangement of samples impact the results of metabarcoding data from an experimental plot in an *in-situ* biomonitoring or environmental impact study context? To determine an optimal sampling strategy at the scale of an experimental plot for metabarcoding, as has similarly been done for classical approaches (Dirilgen et al., 2018; Widenfalk et al., 2016), I implemented a spatially-explicit sampling design to quantify what sampling intensity (numbers of samples and spatial layout) is needed to characterize microarthropod and fungal communities.

2 Methods

2.1 Study Site

Sampling was conducted at the Island Lake Biomass Harvesting Research and Demonstration Area located in the Martel Forest near Chapleau, ON, CA (47° 42' N, 83° 36' W) within the Ontario Shield Ecoregion 3E (Kwiaton et al., 2014). As of 2013, the mean annual temperature and precipitation for Chapleau is 1.7°C and 797 mm, with a growing season of 93 frost free days (approximately June 5-September 6). Dystric Brunisols are the predominant soil great group in the region according to the Canadian System of Soil Classification (Soil Classification Working Group, 1998). Prior to establishment of the experimental area the site was a 50-year-old second growth commercial stand of pure jack pine that did not meet commercial thinning requirements due to low stand density. Therefore, in 2010 the experimental

site was established through collaboration of government (Canadian Forest Service, OMNR-F), industry, and local community stakeholders to determine the impact that intensified biomass harvesting has on boreal forest integrity and sustainability through analyses of changing physical, chemical and biological properties. Harvesting of the site was completed in January of 2011.

2.2 Collection and Identification of Collembola Specimens

A collection of Collembola specimens was contributed by Laurent Rousseau (UQAM) from his thesis work (Rousseau, 2018) to establish a preliminary local reference library. Microarthropods were extracted from bulk soil and moss samples taken from the harvested and control plots at the Island Lake site in 2015 using a standard Tullgren-funnel extraction methodology: soil samples were placed on a wire mesh platform in aluminum funnels and exposed to overhead heating lamps for a period of 5 – 7 days, causing organisms to migrate through the bottom of the funnel where they were collected in vials containing 70% ethanol. Individual specimens were sorted from the extracted community samples with a dissection microscope and morphologically identified using a contrast phase microscope under up to 800x magnification. Identified specimens were stored individually in 0.2 ml tubes containing 70% ethanol.

2.3 DNA Extraction from Collembola Specimens

For molecular identification of the Collembola specimens DNA was extracted from individuals using the TE boiling method described in Aoyama et al. (2015) for rapid, non-

destructive DNA extraction. Specimens were removed from ethanol and placed in 0.2 ml tubes with 20 µl of TE buffer (1 mM Tris-HCl, 0.1 mM EDTA, pH 8.0, Fisher Scientific, Markham, ON, CA). The specimens were then heated in a thermocycler with the following cycle settings: 65°C for 5 min, 96°C for 2 min, 65°C for 4 min, 96°C for 1 min, 65°C for 1 min, 96°C for 30 sec. Fifteen µl of the supernatant containing specimen DNA was then transferred to a 0.2 ml collection tube to be used for PCR.

2.4 PCR of COI Barcodes from Collembola Specimen DNA

To amplify the full-length COI barcode fragment (658 bp) from the DNA extracted from the Collembola specimens, forward and reverse primer cocktails containing two standard COI primers in equal concentrations were used (Hernández-Triana et al., 2014). The forward cocktail (C_LepFolF) consisted of the primers LepF1 (5'-ATT CAA CCA ATC ATA AAG ATA TTG G-3') and LCOI490 (5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3'); the reverse cocktail (C_LepFolR) consisted of the primers LepR1 (5'-TAA ACT TCT GGA TGT CCA AAA AAT CA-3') and HCO2198 (5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3') (Folmer et al., 1994; Hebert et al., 2004). Reactions were performed in 25 µl mixtures containing: 12.5 µl of 2x Phire Green PCR Master Mix (Thermo Fisher Scientific, Waltham, MA, USA), 1.25 µl of forward primer cocktail (10 µM), 1.25 µl of reverse primer cocktail (10 µM), 2.5 µl of undiluted specimen DNA, and 7.5 µl of molecular-grade water (Thermo Fisher Scientific). Reaction cycle conditions were as follows: initial denaturation at 98°C for 30 sec, then 5 cycles of denaturation at 98°C for 10 sec, annealing at 45°C for 10 sec, extension at 72°C for 15 sec, followed by 35 cycles of denaturation at 98°C for 10 sec, annealing at 51°C for 10 sec, extension at 72°C for 15 sec, with a final extension at 72°C for 1 min. 5 µl of reactions were separated via electrophoresis

on a 1% agarose gel to verify reaction success. PCR products were cleaned with the GeneJET PCR purification kit (Thermo Fisher Scientific) and stored at minus 20°C prior to shipment for sequencing.

2.5 Sanger Sequencing and Proofreading of Collembola COI Barcodes

Sanger sequencing of the Collembola specimen barcodes was conducted at the McGill University and Genome Quebec Innovation Centre (Montreal, QC, CA) on a 3730xl DNA analyzer (Applied Biosystems, Foster City, CA, USA). The C_LepFolF and C_LepFolR cocktails were used in bidirectional sequencing reactions with the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems).

Chromatograms from Sanger sequencing were submitted to the PeakTrace Basecaller Online tool (Nucleics, Woollahra, Australia) for processing to increase base-call efficiency. Sequences were then manually edited through visualization of original and PeakTrace-edited chromatograms with the Lasergene SeqMan Pro software (DNASTAR, Madison, WI, USA). Primer residues were removed, and full-length consensus COI barcodes were resolved from bidirectional sequences then exported and saved in FASTA format.

2.6 Identification of Collembola COI Barcodes with the Barcode of Life Database

Before submitting Collembola specimen barcodes for molecular identification, the availability of references for the morphological identification results for each specimen was

assessed. Edited Collembola COI barcode sequences of ≥ 500 bp were then submitted in FASTA format to the Animal Identification Engine of the Barcode of Life Data Systems (BOLD) for taxonomic identification against all COI barcode records (<http://v4.boldsystems.org/>, (Ratnasingham and Hebert, 2007). The lowest level of taxonomic classification provided for the closest match based on percent sequence similarity was recorded. For barcodes with a similarity of $\geq 95\%$, the Barcode Index Number (BIN, (Ratnasingham and Hebert, 2013) was also recorded.

2.7 Sampling and Processing of Soil Samples

A 60 x 82 m plot was established in a Northern section of the site for soil sampling in August 2015 to identify an optimal sampling procedure for metabarcoding of small-scale plots. This area of the site was harvested using a full-tree harvest method and site prepared using disc-trenching and was not replanted after harvest. In disc-trenching a machine is used to scarify the land by using two rotating discs to form furrows (or trenches) and cast the excavated material to the side in loose berms (or mounds). There is a patch of land in between the two discs that is undisturbed by the trenching discs. Three main transects approximately equally spaced apart across the plot were created along the bands of undisturbed soil left by the disc trenching during site preparation. Samples were taken from each transect at 1 m, 5 m, and 10 m intervals with four intermittent samples taken between the transects (Figure 1) for a total number of 44 samples. All samples were collected using corers constructed from 5 cm diameter PVC tubing cut to 10 cm lengths. One end of each corer was tapered to promote movement into soil during sampling. Corers were soaked in 10% bleach for 10 minutes and left to air dry overnight before use. When sampling, woody debris was removed from the soil surface and a rubber mallet was used to drive the entire corer into the soil. Immediately after soil corers and sample were extracted, they were

wrapped in tin foil and sealed in plastic collection bags. Samples were stored in a cooler with freezer packs while in the field and transferred to minus 20°C the same day as possible for long term storage.

Soil cores were thawed at 4°C for 24 – 48 h and processed individually. All sampling material was removed, and soil was manually passed through a stainless steel 2 mm soil sampling sieve (Fieldmaster, Pukekohe, New Zealand) to remove large debris and homogenize the sample. Half of the sample was then transferred to a Ziploc bag and stored at 4°C prior to eDNA extraction, the other half was returned to minus 20°C for long-term storage.

2.8 eDNA Extraction from Soil Samples

eDNA for metabarcoding was extracted from each soil sample in triplicate using the MoBio PowerSoil Kit (MoBio, Carlsbad, CA, USA) with an amended protocol to improve purification efficiency. Approximately 0.3 g of soil was added to the bead tubes and vortexed briefly. 100 µl of $\text{AlK}(\text{SO}_4)_2$ (200 mM, Fisher Scientific) was then added to the bead tubes with the standard C1 solution and this mixture was incubated at 70°C for 10 min (Braid et al., 2003). After loading the entire sample onto the spin filter, prior to the C5 solution wash step, 500 µl of 5.5 M guanidium thiocyanate (GTC, Fisher Scientific) was added to the spin filter and centrifuged at 10000 x g for 1 min (Antony-Babu et al., 2013). The flow-through was then discarded and this step was repeated 3-5 times until the flow-through appeared clear. The kit protocol was then followed for the remaining steps. Triplicate eDNA samples were pooled into a single sample and quantified by spectrophotometry using the Synergy HI microplate reader with a Take3 micro-volume plate (BioTek Instruments, Inc., Winooski, VT, USA).

2.9 PCR Amplification of DNA Barcodes from Soil eDNA

Three common DNA barcodes were selected to target soil arthropods and fungi. For arthropods, two primer sets were used to target the COI-F230 marker that sits at the 5' end of the COI barcode region and the COI-BR5 marker that sits at the 3' end of the of the mitochondrial COI barcode. The F230 region (229 bp) was amplified with the primers LCOI490 (5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3') (Folmer et al., 1994) and 230R (5'-CTT ATR TTR TTT ATI CGI GGR AAI GC-3') (Gibson et al., 2015) and the BR5 region (310 bp) was amplified with the primers B (5'-CCI GAY ATR GCI TTY CCI CG-3') (Hajibabaei et al., 2012) and ArR5 (5'-GTR ATI GCI CCI GCI ARI ACI GG-3') (Gibson et al., 2014). The nuclear ITS2 barcode (240-460 bp) was selected as a target specific to fungi and amplified with the primers fITS9 (5'-GAA CGC AGC RAA IIG YGA-3') and ITS4 (5'-TCC TCC GCT TAT TGA TAT GC-3') (Menkis et al., 2012; White et al., 1990). The v4 region of the nuclear 18S rRNA sequence (384 bp) was used as a broad eukaryotic barcode to capture both arthropod and fungal taxa, and amplified with the primers TAREuk454FWD1 (5'-CCA GCA SCY GCG GTA ATT CC-3') and TAREukREV3 (5'-ACT TTC GTT CTT GAT YRA-3') (Stoeck et al., 2010). All reactions were performed in 25 µl mixtures. For the two regions of the COI barcode, the reaction mixtures included 2.5 µl of 10x KCl PCR buffer (Invitrogen), 1 µl of 50 mM MgCl₂ (Invitrogen), 0.5 µl of 10 mM dNTPs (Invitrogen), 0.5 µl of each primer (10 µM), 0.5 µl of BSA (10 µg/µl, Invitrogen), 0.8 µl of Native Taq polymerase (5 U/µl, Invitrogen), up to 20 ng of purified eDNA, and 16.7 µl of molecular-grade water (Thermo Fisher Scientific). COI reaction cycle conditions were as follows: initial denaturation at 95°C for 5 min, then 35 cycles of denaturation at 95°C for 40 sec, annealing at 46°C for 1 min, extension at 72°C for 30 sec, with a final extension at 72°C

for 2 min. The ITS2 and 18S rRNA barcode reaction mixtures were the same and included 12.5 μ l of 2x Phire Green PCR Master Mix (Thermo Fisher Scientific), 0.5 μ l of each primer (10 μ M), 0.5 μ l of BSA (10 μ g/ μ l, Invitrogen), up to 20 ng of purified eDNA, and 9 μ l of molecular-grade water (Thermo Fisher Scientific). The 18S rRNA reaction cycle conditions were as follows: initial denaturation at 98°C for 3 min, then 5 cycles of denaturation at 98°C for 30 sec, annealing at 57°C for 45 sec, extension at 72°C for 15 sec, followed by 35 cycles of denaturation at 98°C for 30 sec, annealing at 48°C for 45 sec, extension at 72°C for 15 sec, with a final extension at 72°C for 1 min. ITS2 reaction cycle conditions were as follows: initial denaturation at 98°C for 3 min, then 35 cycles of denaturation at 98°C for 30 sec, annealing at 55°C for 60 sec, extension at 72°C for 15 sec, with a final extension at 72°C for 1 min.

All reactions were run once with 35 cycles to verify amplification success before running initial PCRs for producing sequencing amplicons with lower cycle numbers to reduce PCR bias. After this verification, each barcode was amplified from eDNA samples in triplicate with 20 cycle reactions using tagged primers to produce amplicons for sequencing. The forward primer tag was 5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG-3' and reverse primer tag was 5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G-3'. The triplicate reactions were pooled into one sample for purification with the QiaQuick MinElute PCR purification kit (Qiagen, Valencia, CA, USA) and eluted in 10 μ l of molecular-grade water.

2.10 Illumina MiSeq Amplicon Sequencing DNA Barcodes

The purified products for each barcode were quantified with the Quant-iT PicoGreen dsDNA assay kit (Invitrogen) and a protocol adapted for micro-volume quantities using a Take3

micro-volume spectrophotometer/fluorimeter plate (Brescia and Banks, 2010). For each spatial point, the purified and quantified samples were combined in equal amounts to provide a volume of ≥ 10 μ l for sequencing use. These combined samples were then submitted for paired-end sequencing at the Biodiversity Institute of Ontario (Centre for Biodiversity Genomics, University of Guelph, ON, CA) on the MiSeq platform (Illumina Biotechnology Co., San Diego, USA).

2.11 Bioinformatics and Data Analysis

Illumina MiSeq reads were processed using a semi-automated bioinformatics pipeline using a variety of Bash and Perl scripts available at <https://github.com/terrimporter/JesseHoage2018>. Reads from each sample were processed individually. The number of forward and reverse reads were checked for consistency. Reads were paired with SeqPrep using the default settings except that we required a minimum Phred score of 20 at the ends as well as a minimum overlap of at least 25 bp (SeqPrep is available from <https://github.com/jstjohn/SeqPrep>). Cutadapt v1.10 was used to remove the forward and reverse primers using the default settings except that we required a minimum length of 150 bp after trimming primers, required a Phred score of at least 20 at the ends, and allowed no more than 3 ambiguous bases when matching primers (Martin, 2011). VSEARCH v2.4.2 was used to dereplicate the sequences using the default settings with the `--derep_fulllength`, `--sizein`, and `--sizeout` commands (Rognes et al., 2016). USEARCH v9.1.13 was to denoise unique read clusters using the default settings (Edgar, 2016). In this case, denoising refers to the correction of putative sequence errors, removal of contaminant PhiX reads, and removal of putative chimeric sequences. Denoised reads were further clustered into OTUs based on 97% sequence similarity using the uparse pipeline with the `-cluster_otus` command specifying a minimum OTU size of 3

to ensure the removal of any singletons and doubletons (Edgar, 2010). An OTU table was created using the `-usearch_global` command indicating that primer-trimmed reads be mapped to OTUs using a 97% sequence similarity cutoff, considering the plus strand only. For ITS2 sequences, coding regions (ie. leading 18S, internal 5.8S, trailing 28S rRNA) were removed with ITSx (Bengtsson-Palme et al., 2013). At each major step of the pipeline described above, read statistics including read number as well as minimum, maximum, mean, median, and model sequence length were calculated. Taxonomy was then assigned using the Ribosomal Database Project (RDP) Classifier v2.12 (Wang et al., 2007) using a variety of reference sets: 1) the `fungallits_unite` database for ITS2 (Köljag et al., 2013) with an 80% bootstrap support cutoff; 2) a custom 18Sv2.0 classifier based on sequences mined from GenBank [July 2017] available at <https://github.com/terrimporter/18SClassifier> for 18S rRNA with a 70% cutoff at the class rank determined by leave one out testing to result in 99% correct assignments; 3) the COI Eukaryote Classifier v2.0 for the ‘unmodified’ COI reference set using a 30% cutoff at the genus rank determined to result in 99% correct assignments (Porter and Hajibabaei, 2018b); and, 4) the COI Eukaryote Classifier v2.1 for the ‘amended’ reference set with 37 Sanger sequences from local taxa using a 30% cutoff at the genus rank determined to result in 99% correct assignments <https://github.com/terrimporter/JesseHoage2018>.

Merged USEARCH OTU and taxonomic assignment tables were generated and imported into Excel and sequencing summaries were generated directly from OTU table data in Excel. Data were then imported for analysis in R 3.4.0 (R Core Team, 2013) and all figures were generated with the `ggplot2` graphics package (Wickham and Chang, 2016). A two-sample t-test was used to compare the difference in bootstrap support values for Collembola sequences between the unmodified and amended DNA barcode libraries. The `phyloseq` package was used to

format OTU and taxonomy tables for further analysis of the sequence data (McMurdie and Holmes, 2017). To assess the depth of sampling and sequencing, OTU accumulation curves were generated with the `specaccum` function in the `vegan` package using the rarefaction method (Oksanen et al., 2013).

For measuring the impact of spatial distance on community structure the data were first normalized to the smallest library size with the `rarefy_even_depth` function (`phyloseq`) and then transformed to presence/absence values with the `decostand` function (`vegan`). The distance function in the `ecodist` package was then used to generate pairwise compositional dissimilarity matrices using the Jaccard method. For each taxonomic rank, dissimilarity matrices for each corresponding barcode were averaged (i.e. all barcodes were used for the total community; 18S rRNA and ITS2 were used for Fungi; COI and 18S rRNA was used for Arthropoda and Collembola, but COI-BR5 was excluded for Oribatida due to a high proportion of null values). These average dissimilarity matrices were used to characterize the spatial structure of the community through two statistical methods. First, the distance-decay relationship was used to demonstrate the decay of community similarity across space by plotting the average dissimilarity values as a function of the spatial distance between samples and best-fit curves for this relationship were generated with a generalized additive model. Distance-decay relationships have been commonly used in biogeography to describe the spatial structure of communities and identify patterns of dispersal and heterogeneity (Morlon et al., 2008; Soininen et al., 2007). Second, Mantel correlograms were used to quantify the level of spatial autocorrelation in the community and identify patterns of aggregation and dispersal gradients. Spatial autocorrelation is a measure of the influence of a variable on itself across space. In the Mantel correlogram, the Mantel r statistic representing the correlation between 2 dissimilarity matrices is calculated for

spatial distance classes representing fractions of the total distance, with positive values indicating association and negative values indicating avoidance (Legendre and Fortin, 1989; Mantel, 1967). The null hypothesis of the mantel correlogram is that the mean compositional dissimilarity of a distance class is the same as the mean of all other distance classes combined (Dutilleul et al., 2000). To generate the Mantel correlograms from the average dissimilarity matrices, the *pmgram* function (*ecodist*) was used with 20 distance classes and 10 000 permutations (Goslee and Urban, 2017).

3 Results

3.1 Morphological and COI Barcode Identification of Local Collembola Specimens

A total of 61 Collembola specimens were morphologically identified to 26 putative species and processed for COI barcoding. Based on morphological identifications, these specimens represented 10 families and 20 genera within the class Collembola. Thirty-four of these specimens consisting of 17 species yielded COI barcodes of ≥ 500 bp in length and were submitted to the BOLD Identification Engine (Table 1). The database already contained species level references for 5 of the 17 morphological species. For all 34 barcodes, the closest match occurred within Collembola, with a sequence similarity of $\geq 83\%$. Twenty-six barcodes were represented by references with only family level identification. Nineteen of the barcodes yielded matches of $\geq 95\%$ with established BINs, 6 of which contained supplemental records of physical specimens identified to the species level.

At the family level, taxonomic identification was consistent between morphological and barcoding results for 22 of the 34 specimens. Of the 8 barcodes with genus level identification, 7 were consistent with the morphological results, including all 6 specimens with confident matches to species level barcode references in BOLD. Only 2 of the 6 barcodes that were confidently identified by barcoding to a species in the reference database were consistent with the morphological identification results, and the remaining 4 had conflicting identifications.

3.2 Impact of Local Collembola COI Barcodes on OTU Classification Confidence

The distribution of sequences across the range of bootstrap support values was significantly different ($p < 0.0001$) between the amended and unmodified reference libraries (Figure 2). With the unmodified reference library, 35% of the total Collembola reads had bootstrap support values above the recommended confidence limit for accurate identification (≥ 0.3 with using the COI Eukaryote Classifier), and 27% had bootstrap support values of ≥ 0.9 . The addition of 34 locally derived reference barcodes caused this distribution to shift, with 66% and 58% of the total Collembola reads falling into the same two categories, respectively.

3.3 Metabarcoding Sequencing Summary, OTU Distributions and Barcode Specificity

The total number and average number per-sample (5 cm diameter x 10 cm depth) of quality paired sequencing reads and OTUs classified for each barcode and taxonomic rank are summarized in Table 2. Both COI-F230 and 18S rRNA barcodes performed well producing $>10^6$ total quality paired reads each, while COI-BR5 and ITS2 barcodes only produced $>10^5$ and $>10^4$

total reads. The COI-F230 and COI-BR5 barcodes had the highest number of total OTUs as expected with 6706 and 3412, respectively, followed by 1903 OTUs observed with the 18S rRNA barcode and 465 OTUs with ITS2. On a per-sample basis however, 18S rRNA identified a larger number of OTUs than COI-BR5 (493 vs 308), with COI-F230 retaining the highest value (896 OTUs) and ITS2 the lowest value (60 OTUs). Fungi were targeted with ITS2 markers and also detected with the 18Sv4 marker, which identified 465 and 445 OTUs respectively. 18S rRNA resulted in a slightly higher number of fungal OTUs per sample compared to ITS2. Two barcodes were used to target Arthropoda: COI-F230 (4579 OTUs), COI-BR5 (2751 OTUs), and 18S detected Arthropods as well (274 OTUs). COI-F230 captured the highest number of Arthropoda OTUs per sample (303), followed by COI-BR5 (128 OTUs), and 18S rRNA (21 OTUs). COI-F230 far outperformed COI-BR5 and 18S rRNA for Collembola, with the former identifying 67 OTUs and the latter identifying 21 and 13 OTUs respectively. For Oribatida, COI-F230 (54 OTUs) and 18S rRNA (36 OTUs) barcodes performed similarly, while COI-BR5 had minimal success (5 OTUs).

The observed OTU distributions among taxonomic kingdoms varied across the four barcodes (Figure 3). The two COI barcodes shared similar patterns, with metazoan OTUs representing 68.3% and 85.6% of the total OTUs classified for F230 and BR5, respectively. These barcodes also captured a considerable amount of protozoan OTUs, 30.3% for F230 and 10.9% for BR5, and minimal fungal OTUs. The proportions of kingdoms represented by the 18S rRNA barcode was expected to be different than the COI barcodes, and the majority of OTUs were represented as Protozoa (62.2%), with the remaining shared between Fungi (23.4%) and Metazoa (14.4%). The ITS2 barcode only represented fungal OTUs, as expected.

The number of unique genera identified by the 4 barcodes for each taxonomic rank is summarized in Table 3. Comparing the number of fungal genera identified shows that there was very little overlap between the 2 barcodes; only 8.3% of the total 336 fungal genera identified were shared by both 18S rRNA and ITS2, with the remaining genera split evenly between the two. For Collembola, 26% of the total 36 Collembola genera were uniquely identified by the COI barcodes, with only 6 genera discovered solely by 18S rRNA. Contrastingly, 15 and 16 of the 39 total Oribatida genera were identified exclusively by the 18S rRNA and COI barcodes respectively.

3.4 Identification-Sensitivity of Metabarcoding Compared to Traditional Methods

The number of Collembola and Oribatida genera identified through metabarcoding of soil cores ($n = 44$) in this study was compared to the results from a concomitant study at Island Lake by PhD student Laurent Rousseau (UQAM) of the experimental plots at the Island Lake site that implemented traditional methods (Table 4) (Rousseau, 2018). Genus level was used as a reference point for comparison due to the inconsistencies between morphological and molecular identification of the local reference specimens, and because the RDP classifier can only provide 99% confident identification of OTUs at this level. The traditional data set consisted of community extraction results (via Tullgren-funnels) from soil cores ($n = 75$) and moss samples ($n = 15$) collected in 2013 and 2014. Thirty-five Collembola and 29 Oribatida genera were identified in the traditional data set. In comparison, 17 Collembola and 29 Oribatida genera were confidently identified through metabarcoding of soil cores ($n = 44$). There were an additional 19 Collembola and 10 Oribatida tentative genera included in the metabarcoding data that did not

meet the requirements for confident identification. Each method exclusively identified a large proportion of unique genera, as only 14 Collembola and 15 Oribatida genera were observed in both data sets.

3.5 Sampling Depth and Spatial Dependency

The adequacy of our sampling depth was assessed by rarefaction of the number of new OTUs identified per additional sample. Considering the total OTUs identified by all 4 markers (12 288), the average number of samples required to reach >90% of the maximum OTU richness was 10 (Figure 4). This value decreased to 8 and 5 samples for Collembola (Figure 5) and Oribatida (Figure 6), respectively. For each barcode and taxonomic group, the rarefaction curves all plateau before the final sample was reached, indicating that we have saturated our detection of OTUs.

The distribution of physical distances between pairwise samples, grouped into 20 distance classes (~5.1 m within-class range), is represented in Figure 7. Approximately 30% of the 946 pairwise distances were contained within 2 distance classes (40 – 45 and 45 – 50 m). 15 of the 20 distance classes were represented by ≥ 25 pairwise distances and only one distance class had <10 pairwise distances (70 – 75 m).

Distance-decay curves were used to illustrate the trend of increasing community dissimilarity across space. The complete data and taxonomic groups Fungi and Arthropoda showed similar patterns (Figure 8), as the dissimilarity of these groups increased greatly for the first ~25 m before a plateau was reached (total community) or dissimilarity continued to gradually increase (Arthropoda and Fungi) with distance. For Collembola and Oribatida, both

groups differed from the overall Arthropoda trend and differed from each other. For Collembola, the peak dissimilarity was observed between 15-20 m and the trend then decreased from this point until a plateau was reached at >60 m, whereas Oribatida dissimilarity did not peak until ~40 m before gradually decreasing for the remaining measured distance. Also, the ranges of dissimilarity values were different for these two taxonomic groups, with Oribatida having a higher range than Collembola (~0.71-0.77 compared to ~0.61-0.69, respectively).

General comparisons between the taxonomic groups were similar for the distance-decay relationships from the community dissimilarity curves and spatial autocorrelation results from the Mantel correlograms. For the total community, Fungi and Arthropoda subsets there was a similar pattern in test results (Figure 9). All three groups showed significant positive autocorrelation within the first 2 distance classes (~10 m). The Mantel r statistic reached a value of ~0 by the fourth distance class (~15-20 m), after which the trend became random and largely non-significant. Again, Collembola and Oribatida diverged from the trend observed in the broad taxonomic communities and differed from each other (Figure 10). The correlogram for Collembola showed significant autocorrelation for the first distance class but became random and non-significant immediately. Oribatida showed significant autocorrelation for the first 2 distance classes, similar to Arthropoda, but the Mantel r statistic did not reach 0 until the fifth distance class (~20-25 m) before the trend degraded.

4 Discussion

Locally derived barcodes from morphologically identified Collembola specimens had a significant positive impact on the sequence classification confidence of metabarcoding data

suggesting that using molecular and morphological approaches in tandem can improve results. From the study by Rousseau (2018) and the estimates in this study, at least 36 collembolan genera were possibly present at this site in 2015. The specimen barcodes used to supplement our reference library therefore represented approximately one third of the COI generic diversity at the site. With a more intensive effort to barcode and classify individuals from various taxonomic groups, many more sequences from the metabarcoding data could be appended to a classified species to provide further context related to the functional importance of OTUs. For other invertebrate groups (e.g. flying insects and freshwater invertebrates) the barcoding of specimens is a focus of attention and has resulted in greater representation in current databases (Curry et al., 2018; Gibson et al., 2014), however, for soil microarthropods barcoding has not been emphasized.

Based on the rarefaction curves of OTU accumulation with increasing number of samples (Figures 4-6) a minimum number of 10 soil core samples (5 cm diameter x 10 cm depth) is predicted to be adequate for the barcodes selected to obtain ~90% of the total observed OTU richness. The spatial nature of sample collection in a plot could also potentially impact the biodiversity captured. Observations from both the direct comparison of changes in community composition across spatial distance (Figure 8) and statistical analysis with Mantel correlograms (Figures 9, 10) suggested that the community is autocorrelated at distances up to at least 10 m, and samples for metabarcoding should be taken at distances greater than this range to avoid spatial bias in the resulting data. Finally, the choice of DNA barcodes selected can impact the taxonomic range observed in the metabarcoding data. For the two targeted microarthropod groups (Collembola and Oribatida), barcode selection is important to the detection of sub-taxa for each group. The majority of Collembola genera identified originated from the COI barcodes,

but some were only identified using 18S rRNA, whereas half of the Oribatida genera were detected exclusively with either the COI or 18S rRNA barcode. With the results presented in this study, an optimized sampling and metabarcoding approach for biomonitoring soil microarthropod communities in small-scale experimental plots can be recommended (see sections 4.2 and 4.3 below).

4.1 Barcoding Local Specimens Can Improve Global Taxonomy and Ecology

The lack of records in BOLD with genus or species level identifications that corresponded to our morphologically identified specimens hindered our ability to compare the consistency of identification between the methods at this taxonomic level (Table 1). Repeating this comparison in the future as reference databases are updated and merging the BOLD and NCBI databases could improve the sequence identification success and provide a greater number of species level results to compare (Macher et al., 2017; Oliverio et al., 2018; Porter and Hajibabaei, 2018b). With 19 of our specimen barcodes matching to established BINs, only 6 cited a reference species, while taxonomy for the remaining BINs was limited to the family level. However, with the references that were available, the consistency between the identification methods at the family level was 65%, at the genus level was 88%, and at the species level was 29%. Some of these inconsistencies, particularly at the species level, could be attributed to morphological misidentification as physical differences between closely related species can be extremely difficult to discern (Resch et al., 2014). For example, the specimens morphologically identified as *Parisotoma notabilis* and *Xenylla humicola* did not produce barcodes with consistent species level identification, even though both species had representative barcode

references in BOLD. Other specimens that could have been similarly misidentified were morphological species *Isotoma sensibilis* and *Tomocerus flavescens*. However, the inconsistencies that occur at the family level are less likely due to misidentification, as physical differences at this level are more obvious, and could be attributed to DNA contamination from another specimen. For example, morphological species *Isotomiella minor* and *Willemia anophthalma*, which produced barcodes identified to different families and appear to be consistent with the barcodes derived from the morphological specimens *Pygmarrhopalites sp.* and *Folsomia nivalis*.

In a similar approach, Shaw and Benefer (2015) conducted a survey in the United Kingdom to confirm the taxonomic classification of some collembolan specimens and assess the availability of reference material in BOLD for UK Collembola. Of the 48 specimens that were barcoded, 17 were matched to a BIN with consistent taxonomy, 6 were matched to BINs with conflicting taxonomy, and 25 remained unidentified. For the species with conflicting taxonomy, all were consistent at the family level and most at the genus level. The authors cite the likely source for each species level, and include: inaccurate identification of BOLD specimens, which the authors justified based on photographs of source specimens and differences in distinguishable characteristics, coincidental morphology of obscure species, and the likelihood of cryptic diversity within the genus *Sminthurinus*. Notably, 10 of the consistently identified specimens had barcoded references from Canada. Compared to the specimen collection in our study, far less species level references were available than for the collection of UK Collembola. Only one morphological species was shared between the two collections (*Xenylla humicola*), however, our molecular identity was contradictory to this.

BOLD contains a total of 108129 published collembolan records from 77 countries (67% of records from Canada) composing 4864 BINs. Of these records, 48300 (47%) are taxonomically detailed enough to identify 579 species (as of March 20, 2018). It is thus unsurprising that most of our species did not have coexisting references in the database, as the approximately 8700 described species of Collembola (Bellinger et al., 2018) are clearly underrepresented. However, many of these records with limited taxonomy contained photographs of multiple source specimens, demonstrating the difficulty with species identification in Collembola and the scarcity of capable taxonomic expertise. (Bellinger et al., 2018) cites 238 Collembola researchers globally, with 31 located in North America. The focus of taxonomists has traditionally been the description and classification of new species. New taxa are still being discovered on a regular basis, but the rate of discovery relative to the number of Collembola taxonomists has continued to decrease since 1960 (Deharveng, 2004). With DNA barcoding, the prevalence of cryptic diversity is becoming more apparent across many soil invertebrate taxa. Morphological species are commonly found to be polyphyletic and segregated into distinct molecular lineages that can exhibit characteristics more typical of separate biological species, with significant phylogenetic variation and reproductive isolation (Cicconardi et al., 2010; King et al., 2008; Laumann et al., 2007; Martinsson et al., 2015; Porco et al., 2012). Due to this, the predicted species richness derived from geographic distributions and morphological diversity is likely to be significantly underestimated. Based on general ratios of described and unknown species, and additional phylogenetic data, global diversity estimates within Collembola have ranged from 65000 to > 500000 species (Cicconardi et al., 2013; Porco et al., 2014). For taxonomic accuracy, integrative approaches that cumulatively apply detailed and diverse morphological characteristics, multi-marker phylogenetics, and sometimes chemical

data can be applied to confidently distinguish between cryptic species (Heethoff et al., 2011; Schäffer et al., 2010; Sun et al., 2017; Zhang and Deharveng, 2015) and should be a focus for collaboration between taxonomists and molecular ecologists to overcome this taxonomic deficit.

DNA barcodes from morphologically identified specimens have most commonly been used to delineate cryptic species and amend taxonomy or monitor the environment for the presence of specific noteworthy taxa (see previous examples). There have been studies comparing the biodiversity values generated from morphological and metabarcoding approaches (Hamilton et al., 2009; Treonis et al., 2018), however none have quantified the impact that reference libraries amended local specimen barcodes can have on the taxonomic classification of metabarcoding data. In this study, the addition of a small collection reference barcodes derived from local specimens significantly increased the quantity of metabarcoding data with confident taxonomic classification (Figure 2). As the diversity of Collembola is vastly underrepresented in molecular databases, it is possible that the addition of any new reference sequences to a library could have this effect. However, there is evidence that molecular species/lineages could be the result of geographic distributions (Cicconardi et al., 2010; Shaw et al., 2013), and comprehensive reference libraries from broad geographic locations could help to elucidate the significance of geography for speciation in Collembola. Regardless, this shows the impact that the taxonomic deficit can have on metabarcoding studies. In a recent survey of the availability of COI references in GenBank and BOLD for 2534 North American freshwater invertebrate genera, Curry et al. (2018) found a large variance in the representation among taxa. In total, 61.2% of genera were represented by at least one associated reference sequence, however, this ranged from 73.9% for Mollusca to 15.3% for Nematoda. In the context of soil biodiversity, although no similar assessment has been conducted, it is reasonable to hypothesize that the taxonomic deficit

could be significantly greater due to the added obstacles associated with soil taxonomic research. In effort to reduce this deficit and increase the power of metabarcoding data, the development of local reference libraries should be valued as a key step in metabarcoding studies.

To predict the ecological impacts that can result from changes to a community, knowledge of the functional contribution of the community members in the ecosystem studied is required. For soil microarthropods, most of this knowledge is derived from the morphological study of specimens and observations from feeding and behavioral experiments. There is a current focus on quantifying trait diversity in communities in comparison to taxonomic diversity for predicting the functional changes of the community and the ecological consequences of these changes (e.g. Farská et al., 2014; Lindo et al., 2012; Sechi et al., 2017; Widenfalk et al., 2016). For molecular data, a trait-based approach is limited by the number of OTUs that can be identified to a species that has been characterized. As soil microarthropod diversity is currently underrepresented in barcode reference libraries, a large amount of the biodiversity data generated would be excluded with a trait-based approach. The value of OTUs or other forms of molecular classification is the ability to comparing communities between treatments and with reference control sites. Differences in communities between these types of conditions can be quantified and can include some diversity that is not observed by traditional methods (Cowart et al., 2015; Emilson et al., 2017; Gibson et al., 2015; Treonis et al., 2018). Also, molecular results can be used to direct specimen collection and barcoding initiatives to areas with high levels of unknown OTUs or target specific OTUs with a strong response to a treatment or that are predicted to have an important ecological role. Applied cooperatively, molecular and traditional can be used to improve the overall performance of biomonitoring studies and further our understanding of soil biodiversity.

4.2 Phylogenetic Barcode Performance Varies Across Taxa

There was a large variance in the number of quality paired reads and OTUs generated by each marker (Table 2). The COI-BR5 and ITS2 sequencing results indicate that there was poor amplification in the PCR step, leading to a much lower number of reads generated compared to the COI-F230 and 18S rRNA barcodes. Based on prior studies using these COI primer sets, the F230 and BR5 barcodes were expected to produce similar quantities of reads and OTUs (Gibson et al., 2014, 2015). The COI barcodes did provide similar relative abundances of OTUs across major taxa as expected (Figure 3), however, there was a fold-difference observed between the number of OTUs identified by the two barcodes, resulting in poor representation of Collembola and Oribatida by BR5 (Table 3). The high relative abundance of protozoan OTUs identified with the 18S rRNA barcode was expected and could be a useful addition in ecological analyses, as Protozoa contains many ecologically relevant taxa (Geisen et al., 2018), however this group was not the target here. Variance in the number of arthropod OTUs identified by the COI barcodes compared to 18S rRNA was likely due to inherent differences between the genes as mtDNA mutates more rapidly than rDNA, resulting in a higher level of sequence variability (Knowlton and Weigt, 1998). Due to this, COI has been shown to provide much higher estimates of faunal OTU richness than 18S rRNA (Tang et al., 2012). Similarly, ITS2 has been shown to provide higher estimates of fungal OTU richness and cover a broader taxonomic range than 18S rRNA (Schoch et al., 2012; Tedersoo et al., 2015). Since there were several arthropod and fungal taxa exclusively identified by the 18S rRNA barcode, the final OTU richness could have been underestimated. Using COI primers specific for Collembola and Oribatida could promote the identification of some of the taxa exclusive to the 18S rRNA barcode and provide a more

accurate OTU richness estimate (Deagle et al., 2014). Although using a 3% OTU clustering threshold for COI could be considered low for OTU delimitation, resulting in an overestimate of OTU richness. In other studies, the average sequence divergence was found to be ca. 11% between animal species pairs (Hebert et al., 2003b), however even with a conservative threshold (e.g. 14% for Collembola in Porco et al., 2014) OTU richness consistently been found to be higher than estimates based on morphological results. Considering the results from the barcode markers used here, it is recommended to use both COI and 18S rRNA markers for metabarcoding soil microarthropods to more accurately estimate diversity.

The soil microarthropod communities at the Island Lake experimental site have been recently studied by Rousseau (2018) using traditional methods, providing a comparative measure of local biodiversity for the metabarcoding data generated here. Considering the confidence limits for OTU identification in the molecular data, there was an equal number of Oribatida genera identified between the two methods, and for Collembola, the number of genera identified in the morphological data set was more than double that in the molecular data set (Table 4). Less than half of the genera identified in the molecular data set were also identified in the morphological data, and the majority of genera were exclusively observed with only one of the methods. The sequence classifier can be considered conservative, as it is optimized to favor false negative assignments to prevent false positives, and the number of genera with accurate classification could be greater than indicated (Porter and Hajibabaei, 2018b). Disregarding the identification confidence limit of the metabarcoding data, the potential total number of Collembola genera identified between the two data sets was similar, but a greater number of Oribatida genera was identified through metabarcoding (Table 4). By expanding the local reference library through further collection and barcoding of local specimens, a greater amount

of OTU taxonomic identities could be confidently resolved and provide a more accurate estimate of richness. In contrast, because the traditional data set surveyed the experimental treatment and control plots, some of the unique genera observed could be attributed to the different environmental conditions in these plots compared to the spatial sampling plot used for the metabarcoding data set, or simply due to the spatial heterogeneity of this community. Most notably, moss and soil samples from the unharvested control plots yielded many taxa absent from the harvested plots (Rousseau et al., 2018). The moss layer, or “bryosphere”, which was only present in control plots, is an important habitat for many microarthropod groups and supports greater soil richness and diversity (Lindo and Gonzalez, 2010). Including moss samples from control plots could have reduced the number of taxa uniquely discovered in the traditional data set.

Previous studies comparing community data sets derived from traditional and molecular methods have also found differences between the approaches. Porco et al. (2014) conducted a survey of Collembola diversity across a sub-Arctic region of Canada. Specimens were collected with a variety of methods, identified to morphological species with several representatives from each species barcoded. The richness in the area was estimated to be >50% higher based on the barcoding data, even with a conservative 14% threshold for OTU clustering. This greater richness was attributed to 7 of the 45 morphospecies appearing as species complexes through the barcoding data, each containing several genetically distinct lineages. In another context, molecular and traditional community datasets have been used to show to present similar overall trends across environments or experimental treatments, however significant differences within results are observed. Hamilton et al. (2009) is one example of this, where molecular and traditional data sets from a hardwood forest and a grassland pasture were compared. The general

trend between environments of an overall high abundance of nematodes and Oribatida and a greater arthropod to nematode ratio in the forests was consistent between methods. However, relative abundances of all taxa did not correlate between the methods and some taxa (e.g. Tardigrada) were only observed in the molecular data set. In a survey of macroinvertebrates boreal watersheds, Emilson et al. (2017) compared the ability of metabarcoding and morphological results for detecting gradients in the stream environment. Using common metrics for aquatic macroinvertebrate biomonitoring, the two methods were positively correlated and identified similar relative gradients across streams, however the metabarcoding approach provided higher estimates of generic and species/OTU richness in the watersheds. In another example, Treonis et al. (2018) compared differences within the nematode community across 3 agricultural cropping systems. There was a similar correlation between nematode community structure soil environmental properties between the molecular and traditional data sets, but significant differences in the relative abundances of some nematode families were observed. Again, there were groups exclusively observed with only one of the methods. In all these examples, the biases presented by both traditional and molecular methods make it difficult to identify which results in a more accurate estimation of the soil community and direct comparisons between the techniques can be deceptive as each is measuring different values (individual specimens versus amplified DNA barcodes). Cooperatively resolving these biases through the continued collection of barcoded local voucher specimens and further development and testing of primer sets for the targeted taxonomic groups will help to continue improving the quality of molecular data and the accuracy and depth of taxonomic reference databases.

4.3 Number of Samples and Spatial Arrangement Can Impact Metabarcoding Results

The number of samples required for accurate metabarcoding of soil microarthropods in small-scale plots had not been quantified previously. In the spatial plot established at Island Lake, 90% of the total OTUs observed were discovered in an average of 10 samples (Figure 4), with an even lower number of samples required to reach this threshold for the targeted taxonomic groups Collembola and Oribatida (Figures 5 and 6). Based on these results, it is recommended that experimental plots of similar sizes should be sampled with 10 soil cores for metabarcoding to confidently detect at least 90% of the microarthropod community. Additional sampling would incur additional cost without much statistical gain. These results conflict with those reported in other studies, where the number of samples required to reach 90% of the observed Collembola or Oribatida richness is consistently much greater. Species accumulation curves based on traditional sampling methods at spatial scales similar to this study rarely reach a plateau with greater sample numbers (up to 370 samples in Ponge and Salmon, 2013) and often require more than 20 samples to obtain 90% of the estimated species richness (Dirilgen et al., 2018; Minor, 2011; Widenfalk et al., 2016). The discrepancy between the results reported here and in previous studies could be due to the greater volume of data that is generated from HTS than from traditional morphological methods. Traditional data sets are typically generated from 10^4 - 10^5 individuals (e.g. Lindo and N. Winchester, 2008; Porco et al., 2014; Rousseau et al., 2018; Widenfalk et al., 2016) whereas molecular data sets can contain $>10^6$ sequences. With greater sampling depth in each sample, this could reduce the number of samples per plot required to accurately estimate richness. Although with metabarcoding, the resulting sequences and OTUs are limited to the products of PCR which can be selective to some members of the community

over others due to primer bias, careful primer selection, optimization of PCR reaction conditions, and replication within samples can limit this bias (Ficetola et al., 2015; Schmidt et al., 2013).

The optimal spatial arrangement of soil samples for metabarcoding soil microarthropods had not been previously demonstrated. The comparison of pairwise community dissimilarity across distance and results from the Mantel test for spatial influence of community dissimilarity demonstrated that there was a spatial trend in the dissimilarity of the complete soil community with significant autocorrelation at distances up to 10 m (Figures 8 and 9). The spatial trends for Collembola and Oribatida both diverge from the trend observed in the total community in different aspects (Figure 10). Collembola appeared to have a shorter range of spatial dependency than the total community, as the dissimilarity curve peaked at a smaller distance and significant autocorrelation was only calculated for the first distance class (5 m) of the Mantel correlogram. For Oribatida, an opposite trend was observed, as the dissimilarity curve peaked at a larger distance that the total community and the Mantel statistic remained positive until the fourth distance class, instead of the third as observed in the total community analysis, although only the first two distance classes (10 m) maintained significant values. For both groups, the correlogram shows an oscillating trend of peaks and valleys at larger distances, which is typical of an aggregated community and common for soil microarthropods (Dirilgen et al., 2018). Collembola are considered to be more efficient colonizers of soil than Oribatida due to their higher fecundity and rates of dispersal, and shorter ranges of autocorrelation were expected and have been observed previously, however, the ranges of spatial autocorrelation in mature forest soils have been observed at 0.5-1.5 m for Collembola (Widenfalk et al., 2016) and up to 40 m for Oribatida (Minor, 2011). Differences associated with using a metabarcoding approach, as previously outlined, or due to the impact of recent tree harvesting at the Island Lake site could cause the

differences in autocorrelation distances observed here compared to the literature. With the recent perturbation of the soil and elimination of soil cover in this study, it is possible that the microarthropod community is still unstable and the patterns of competition and spatial partitioning that occur in a mature forest soil has not developed yet. Microarthropod community composition at small-scales have been shown to be heterogeneous and explained by spatial variables more than environmental variables (Dirilgen et al., 2018; Ponge and Salmon, 2013; Widenfalk et al., 2016), although species have been shown to aggregate in similar patterns due to both variables depending on behavioral differences (Widenfalk et al., 2018). At this site and in previous studies, Collembola have demonstrated greater seasonal and annual variability, and contrasting responses to environmental disruption than Oribatida (Malmström et al., 2009; Rousseau et al., 2018). These observations are results of the general differences related to their r- and K-selection strategies and could be the source of separation in the autocorrelation results for the two groups here. Future work that cooperatively applies molecular and traditional methods to amend identifiable taxa from molecular biodiversity data with traits like fecundity and dispersal rates could result in a more detailed understanding of the differences observed in spatial community structure.

5 Conclusions

Soil microarthropods have the potential to be used as powerful indicators of soil quality, however their use in biomonitoring applications has been limited due to inefficiencies with traditional methods of sample collection and specimen identification. With the adoption of modern molecular methods like DNA barcoding and metabarcoding, these ecologically significant communities can become more accessible for biomonitoring applications. Demonstrated here, DNA barcoding can be used to supplement traditional and molecular

taxonomic classification, and metabarcoding can be optimized for plot-level microarthropod community detection through barcode target selection and sampling design. First, applying traditional and molecular methods in a coordinated approach is beneficial for increasing the knowledge of microarthropod biodiversity and improving the quality of metabarcoding data. The collection of morphologically identified and barcoded local specimens included many taxa that were not previously catalogued in molecular reference libraries and the amended barcode database significantly improved the identification efficiency of metabarcoding generated sequences. Second, barcode selection impacts the observed communities and results can vary between microarthropod groups. For both Collembola and Oribatida, some genera were only identified with one of the three barcodes used to target microarthropods. Also, with the addition of one additional barcode, the fungal community was also assessed from the same samples used for microarthropods. Finally, an optimal sample quantity and arrangement for small-scale plots was recommended through the analysis of sampling depth and spatial dependency of the microarthropod community. To obtain 90% of the observed community, 10 samples with >10m separation between samples is recommended to reduce the impact of spatial autocorrelation on the observed community. With the large taxonomic deficit in soil biodiversity, tandem approaches that apply both traditional and modern methods in a complementary manner should be a focus of research moving forward to catalogue biodiversity and generate a greater understanding of soil ecology. With this, the ecological impacts of land development, bioremediation and reclamation, and climate change could become clear, and sustainable land management practices could become more readily identified.

Table 1: Morphological species and top match in BOLD based on COI barcode identification of local Collembola specimens. Representation of morphological species in BOLD is given as +/- . BINs are provided for COI barcodes with $\geq 95\%$ similarity to top match.

Morphological Identification (Order: Family: Species)	BOLD Reference for Species	COI Barcode Identification (BOLD top match)
Entomobryomorpha: Entomobryidae: <i>Entomobrya confusa</i>	-	Entomobryidae, BOLD:ACK5136 (>99%)
Entomobryomorpha: Entomobryidae: <i>Lepidocyrtus curvicolis</i>	-	Entomobryidae (>89%) Entomobryidae, BOLD:ACT8552 (>99%)
Entomobryomorpha: Entomobryidae: <i>Lepidocyrtus helenae</i>	-	Entomobryidae, BOLD:ACM2038 (>96%)
Entomobryomorpha: Entomobryidae: <i>Lepidocyrtus unifasciatus</i>	-	Entomobryidae, BOLD:ACT8552 (>99%)
Entomobryomorpha: Entomobryidae: <i>Orchesella imitari</i>	-	Entomobryidae, BOLD:ABA5351 (>97%) Entomobryidae, BOLD:ABA5351 (>98%)
Entomobryomorpha: Isotomidae: <i>Folsomia nivalis</i>	+	Hypogastruridae (>83%) <i>Folsomia nivalis</i> , BOLD:AAI2059 (100%)
Entomobryomorpha: Isotomidae: <i>Isotoma sensibilis</i>	-	<i>Isotoma riparia</i> , BOLD:AAA5260 (>98%) <i>Isotoma riparia</i> , BOLD:AAA5260 (>98%)
Entomobryomorpha: Isotomidae: <i>Isotomiella minor</i>	+	Arrhopalitidae, BOLD:ACB2244 (>99%)
Entomobryomorpha: Isotomidae: <i>Metisotoma grandiceps</i>	-	Neanuridae (>87%) Entomobryidae (>84%)
Entomobryomorpha: Isotomidae: <i>Parisotoma ekmani</i>	+	<i>Parisotoma ekmani</i> , BOLD:AAI2638 (100%)
Entomobryomorpha: Isotomidae: <i>Parisotoma notabilis</i>	+	Isotomidae, BOLD:ACO0583 (>98%) <i>Parisotoma ekmani</i> (>86%) Neanuridae (>88%)
Entomobryomorpha: Tomoceridae: <i>Tomocerus flavescens</i>	-	<i>Tomocerus sp.</i> , BOLD:AAA7969 (>99%) <i>Tomocerus sibiricus</i> , BOLD:AAA7968 (>99%)
Poduromorpha: Hypogastruridae: <i>Willemia anophthalma</i>	-	Onychiuridae (>84%) <i>Folsomia nivalis</i> , BOLD:AAI2059 (>99%) Onychiuridae (>84%)
Poduromorpha: Hypogastruridae: <i>Xenylla humicola</i>	+	Hypogastruridae (>83%) Hypogastruridae (>83%)
Poduromorpha: Onychiuridae: <i>Onychiurus subtenuis/similis</i>	-	Onychiuridae (>85%) Onychiuridae (>85%)
Symphyleona: Arrhopalitidae: <i>Pygmarrhopalites sp.</i>		Hypogastruridae (>83%) Sminthuridae, BOLD:ACQ8656 (>97%)
Symphyleona: Arrhopalitidae: <i>Pygmarrhopalites benitus</i>	-	Arrhopalitidae, BOLD:ACB2244 (>99%) Arrhopalitidae (>85%) Sminthuridae, BOLD:ACQ8656 (>97%) Sminthuridae, BOLD:ACQ8656 (>97%) Arrhopalitidae (>93%)

Table 2: Sequencing summary of the 4 DNA barcodes. Values are represented as all 44 samples cumulatively (total) and the average per sample. Read numbers are for paired and quality filtered results.

	COI-F230		COI-BR5		18S		ITS2	
	Total	Average	Total	Average	Total	Average	Total	Average
Reads	1.44x10 ⁶	3.28x10 ⁴	1.68x10 ⁵	3.82x10 ³	2.77x10 ⁶	6.29x10 ⁴	7.27x10 ⁴	1.65x10 ³
Total OTUs	6706	896	3214	308	1903	493	465	60
Fungi OTUs	-	-	-	-	445	101	465	60
Arthropoda OTUs	2477	303	1399	128	103	21	-	-
Collembola OTUs	67	16	21	3	13	3	-	-
Oribatida OTUs	54	6	5	1	36	11	-	-

Table 3: Comparison of unique genera identified by each barcode for the three targeted taxonomic groups. Collembola and Oribatida are represented by the barcodes COI-F230, COI-BR5, and 18S rRNA. Fungi are represented by the 18S rRNA and ITS2 barcodes

Taxonomic Group	Total Genera	COI			18S		ITS2	
		Unique Genera (F230, BR5)	F230 OTUs	BR5 OTUs	Unique Genera	OTUs	Unique Genera	OTUs
Fungi	336	-	-	-	156	445	152	465
Collembola	36	26 (14, 3)	67	21	6	13	-	-
Oribatida	39	16 (11, 1)	54	5	15	36	-	-

Table 4: Comparison of Collembola and Oribatida genera identified from Island Lake with a traditional approach from Rousseau et al. (2018) and through metabarcoding. We used the minimum COI bootstrap proportion (BP) cutoffs suggested by Porter and Hajibabaei (2018b). Bracketed values indicate genera shared with morphological data set. Values in parentheses indicate number of taxa unique to metabarcoding data sets.

Taxonomic Group	Morphological Data	Metabarcoding Data with confidence limit	Total Metabarcoding Data
Collembola	35	17 (12)	36 (14)
Oribatida	29	29 (10)	39 (15)



Figure 1: Schematic diagram for spatial sampling of exploratory plot at Island Lake site. All distances are in metres.

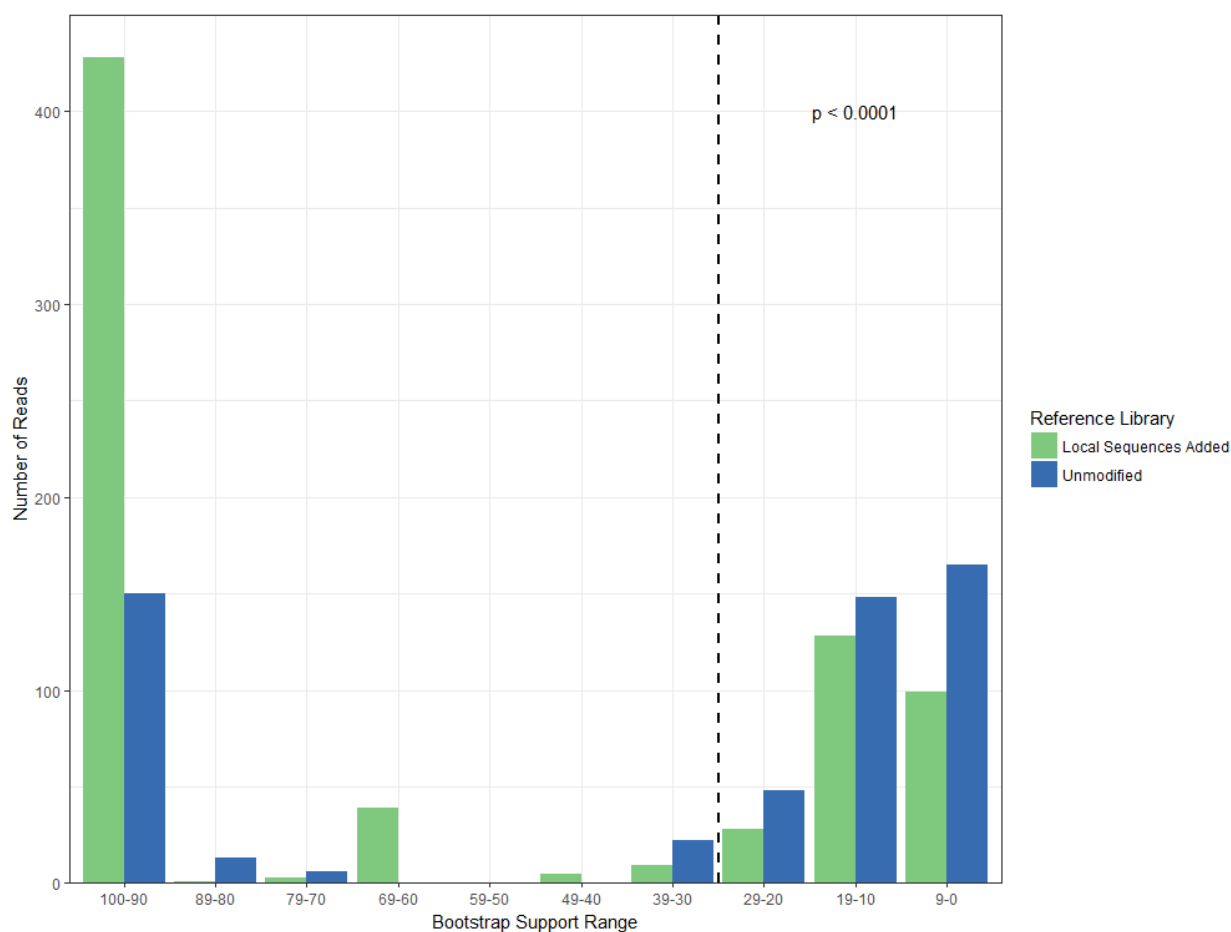


Figure 2: Comparison of bootstrap support values for taxonomic identification of Collembola reads between the unmodified reference library and one supplemented with the local Collembola COI barcodes. The vertical line indicates the recommended bootstrap support cutoff (30%) for 99% confident sequence identification.

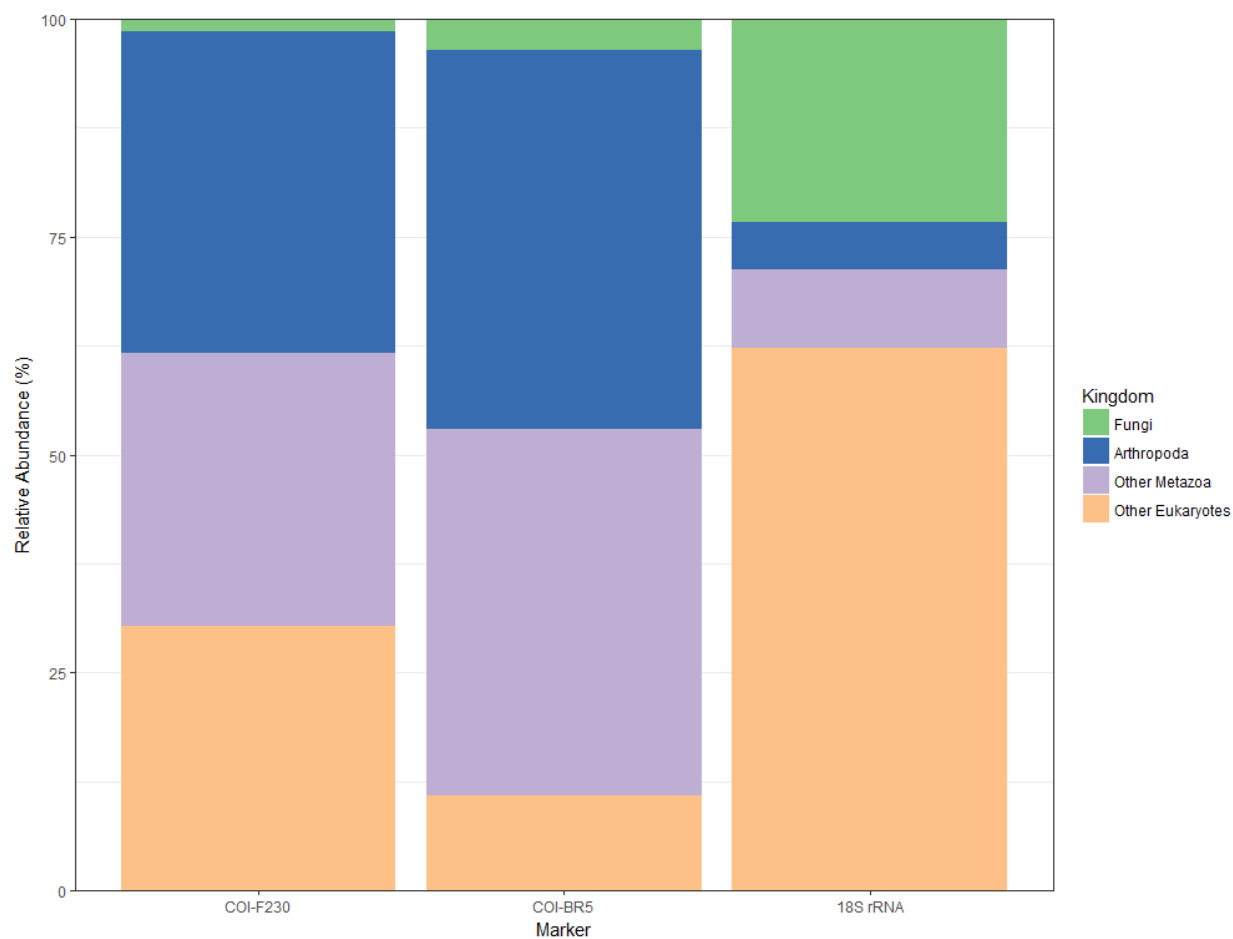


Figure 3: Distribution of the relative number of OTUs classified for each barcode, separated by taxonomic kingdom. ITS2 was not included as all OTUs belonged to Fungi.

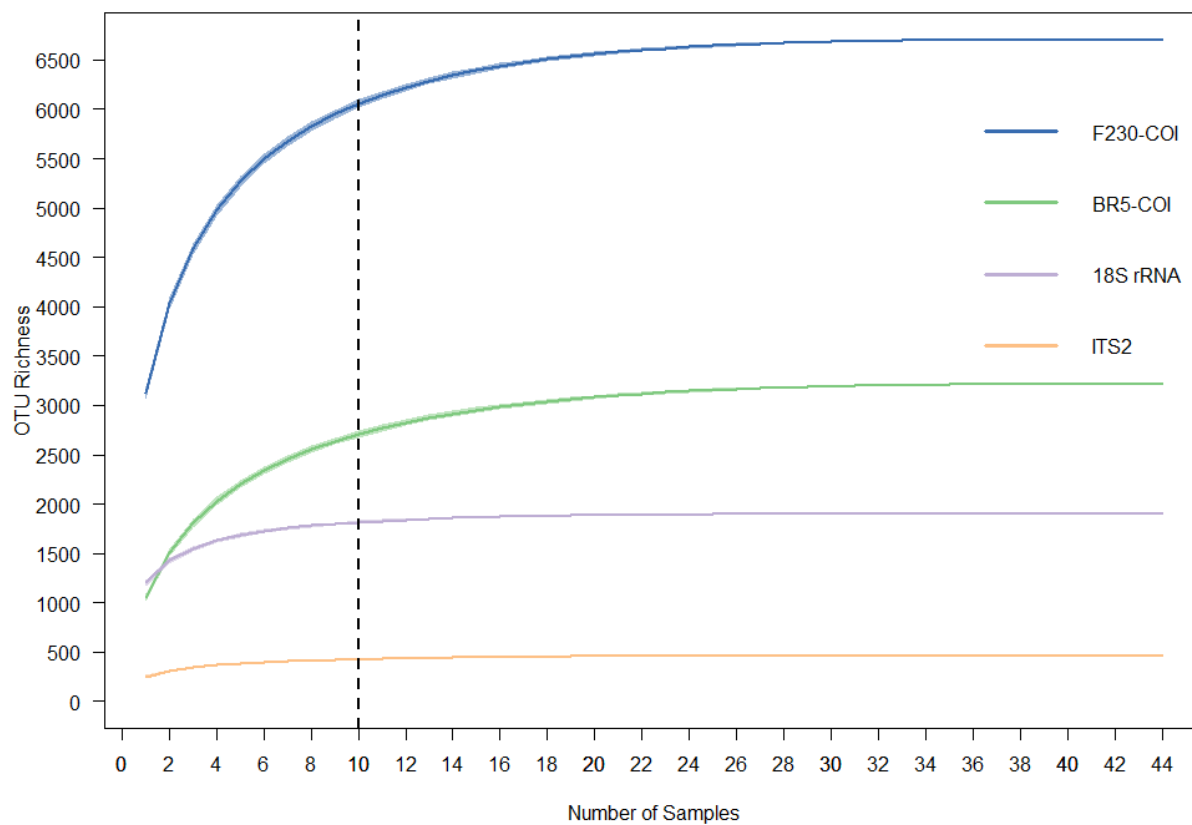


Figure 4: Rarefaction curve with 95% confidence limits for the accumulation of total OTU richness with increasing number of samples for each barcode. Vertical line indicates the average number of samples at which $\geq 90\%$ of species richness is achieved.

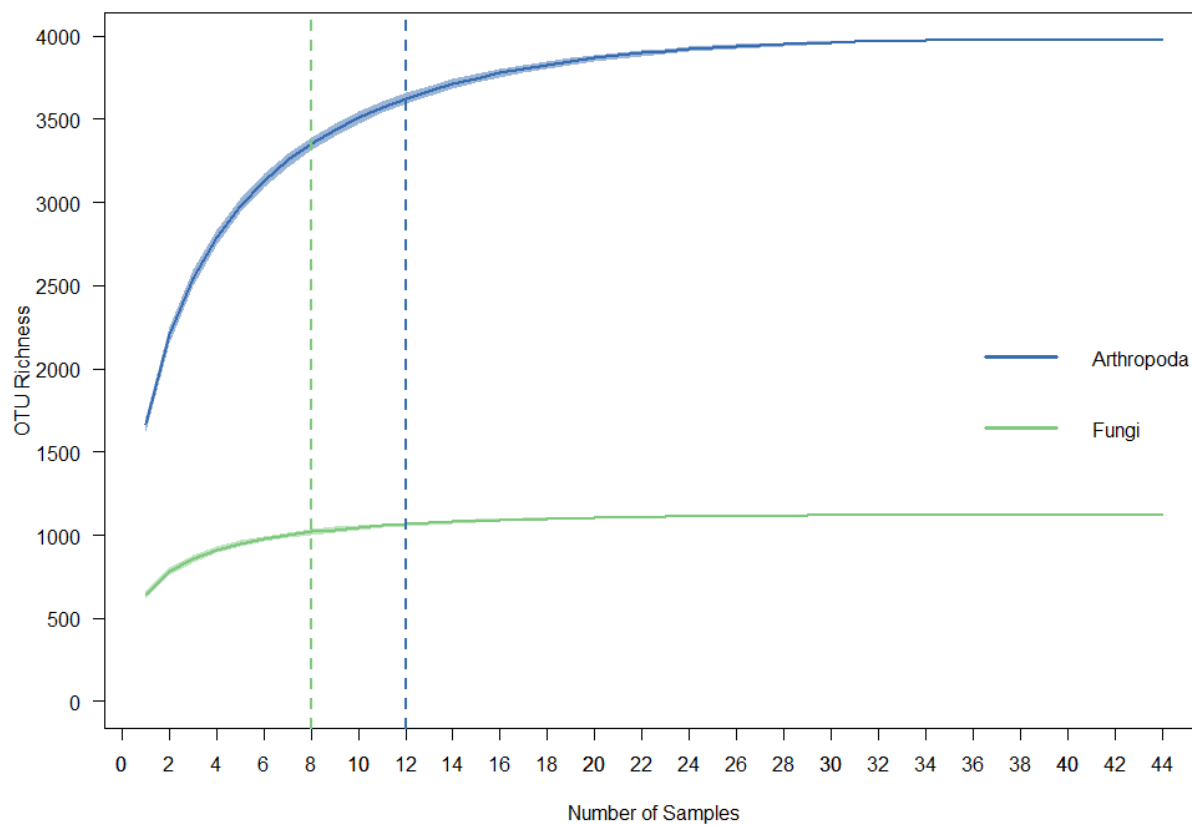


Figure 5: Rarefaction curve with 95% confidence limits for the accumulation of total OTU richness with increasing number of samples for Arthropoda and Fungi OTUs. Vertical lines indicates the number of samples at which $\geq 90\%$ of species richness is achieved.

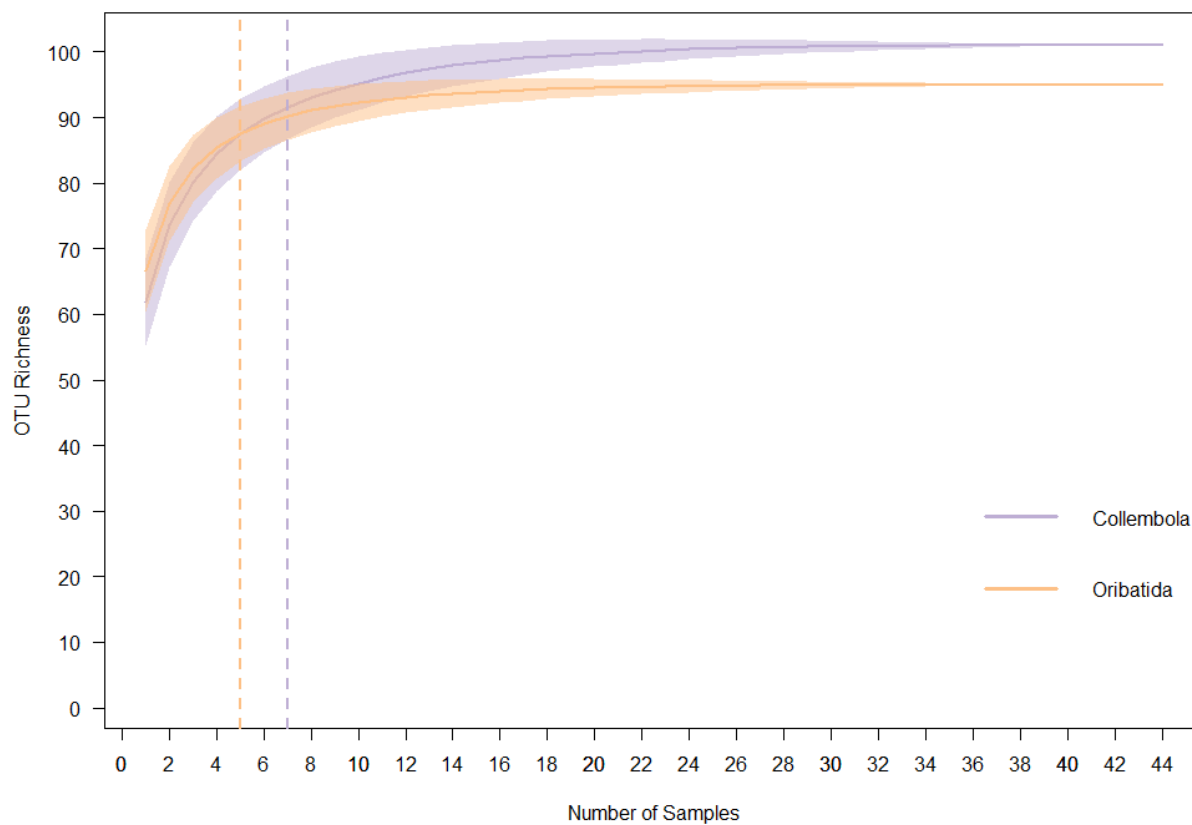


Figure 6: Rarefaction curve with 95% confidence limits for the accumulation of total OTU richness with increasing number of samples for Collembola and Oribatida OTUs. Vertical lines indicates the number of samples at which $\geq 90\%$ of species richness is achieved..

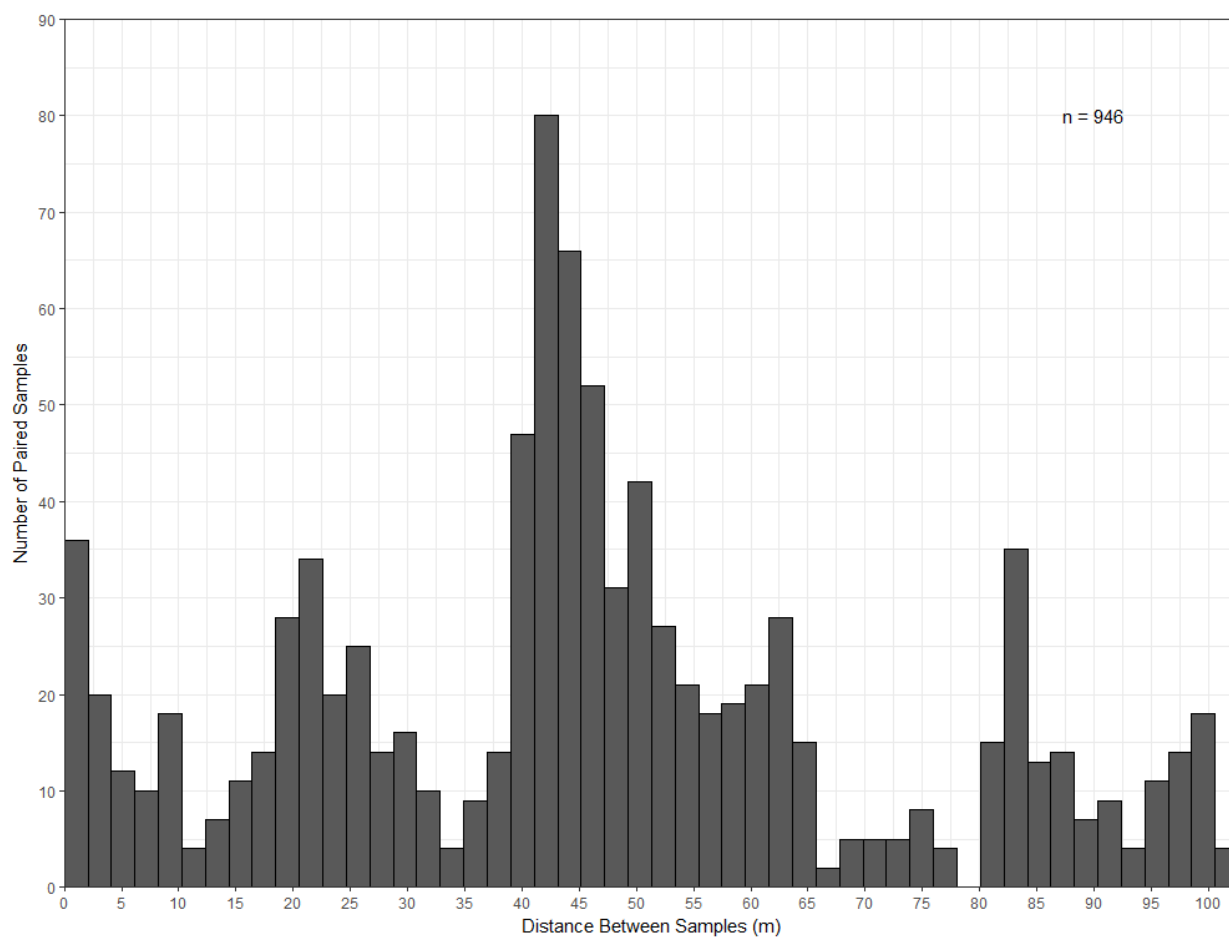


Figure 7: Histogram of all spatial distances between pairwise samples from exploratory plot, summarized into 2 m intervals.

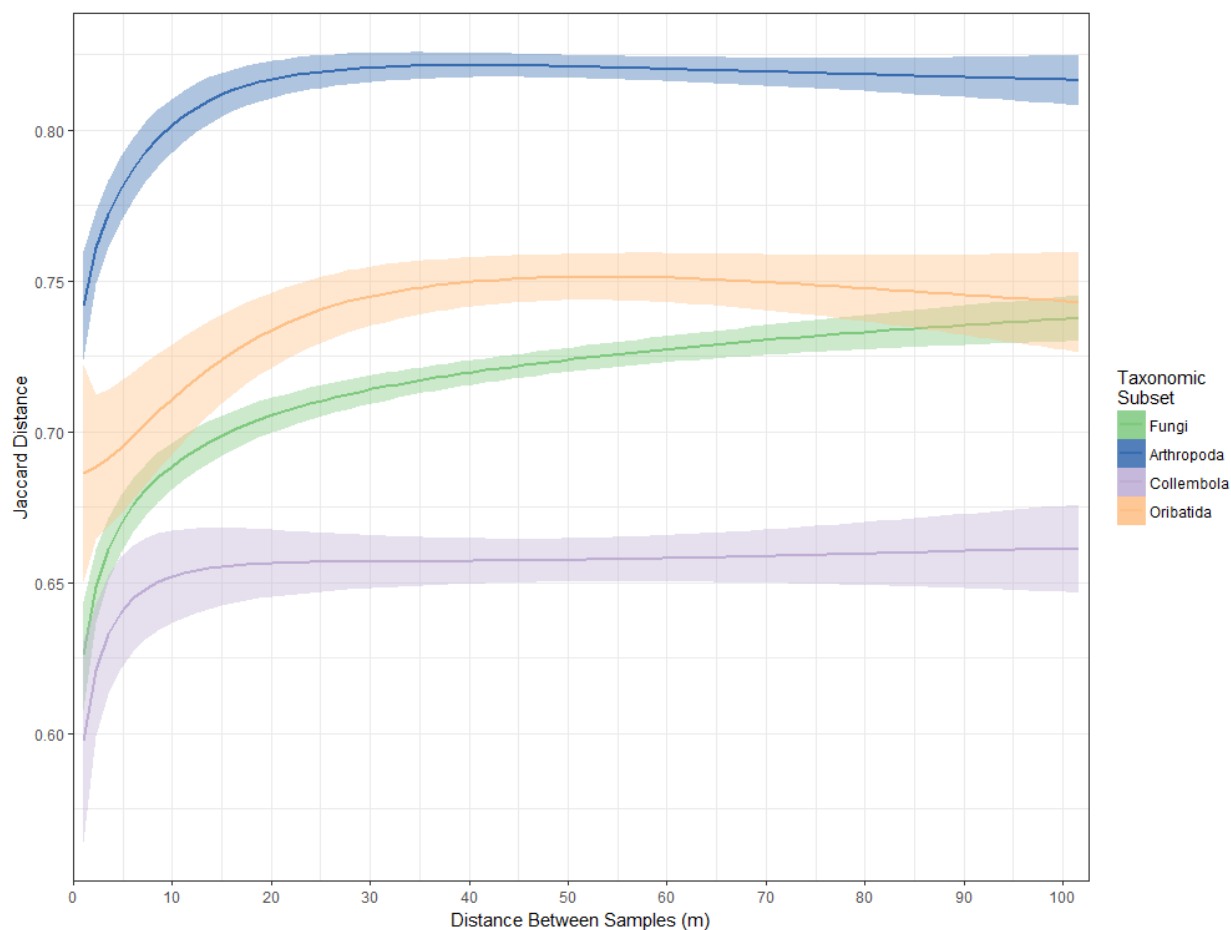


Figure 8: Relationship between Jaccard distance given spatial distance for each taxonomic rank with 95% confidence limits. Curve calculated with generalized additive model.

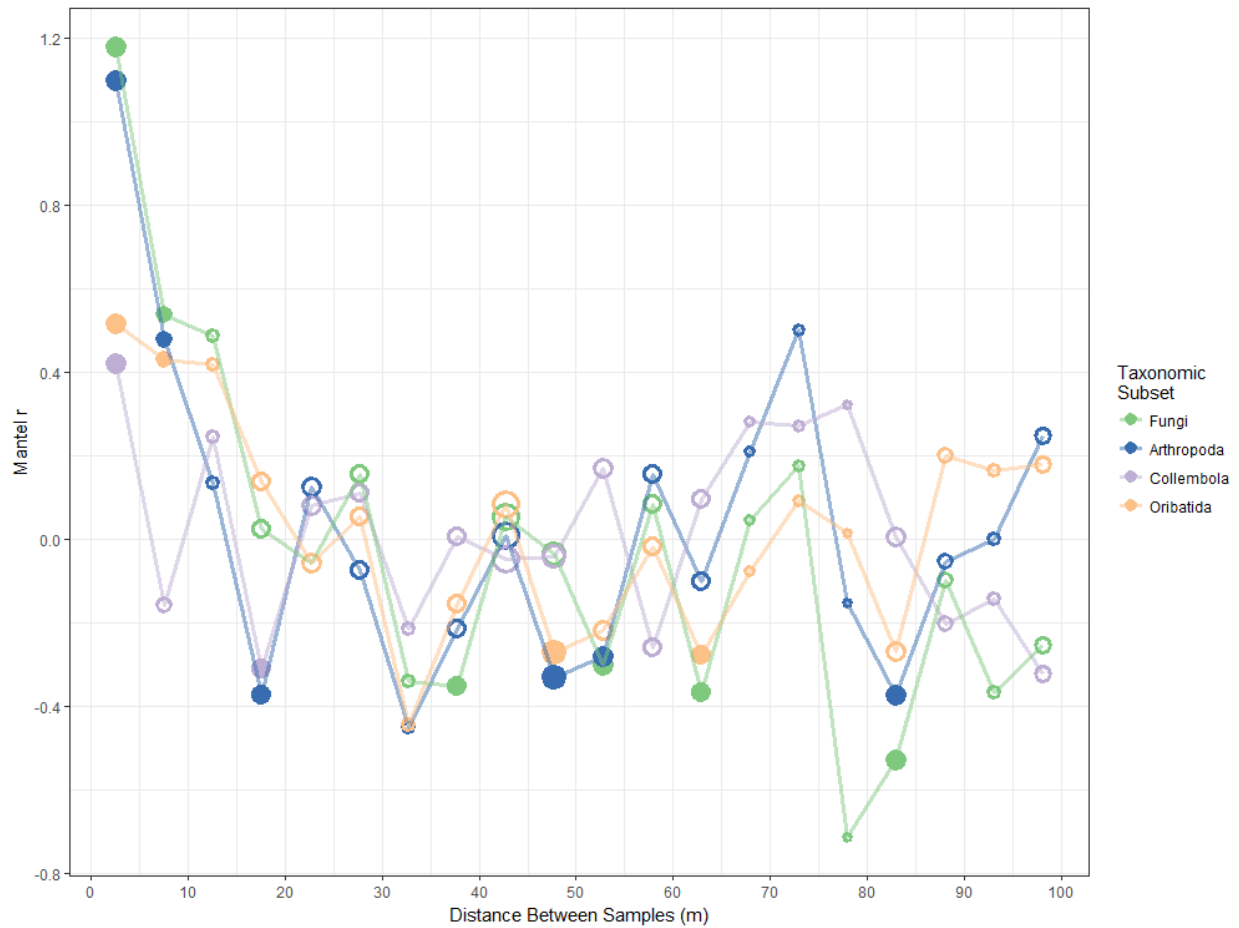


Figure 9: Correlogram of Mantel r values of 20 distance classes for each taxonomic rank. Mantel r values calculated from community dissimilarity and spatial distances. Solid points indicate significant p -values (≤ 0.05) and the size of points indicate relative number of values within each distance class.

References

- A'Bear, A.D., Jones, T.H., Boddy, L., 2014. Size matters: What have we learnt from microcosm studies of decomposer fungus–invertebrate interactions? *Soil Biol. Biochem.* 78, 274–283. <https://doi.org/10.1016/j.soilbio.2014.08.009>
- Achat, D.L., Fortin, M., Landmann, G., Ringeval, B., Augusto, L., 2015. Forest soil carbon is threatened by intensive biomass harvesting. *Sci. Rep.* 5, 15991. <https://doi.org/10.1038/srep15991>
- Addison, J.A., Barber, K.N., 1997. Response of soil invertebrates to clearcutting and partial cutting in a boreal mixedwood forest in Northern Ontario (No. GLC-X-1). Great Lakes Forestry Centre, Canadian Forest Service (Natural Resources Canada), Sault Ste. Marie, ON.
- Addison, J.A., Trofymow, J.A., Marshall, V.G., 2003. Functional role of Collembola in successional coastal temperate forests on Vancouver Island, Canada. *Appl. Soil Ecol.* 24, 247–261. [https://doi.org/10.1016/S0929-1393\(03\)00089-1](https://doi.org/10.1016/S0929-1393(03)00089-1)
- Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., Gnirke, A., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* 12, R18. <https://doi.org/10.1186/gb-2011-12-2-r18>
- André, H.M., 2002. Soil biodiversity: myth, reality, or conning? *Oikos* 96, 3–24. <https://doi.org/10.1034/j.1600-0706.2002.11216.x>
- Anslan, S., Tedersoo, L., 2015. Performance of cytochrome c oxidase subunit I (COI), ribosomal DNA Large Subunit (LSU) and Internal Transcribed Spacer 2 (ITS2) in DNA barcoding of Collembola. *Eur. J. Soil Biol.* 69, 1–7. <https://doi.org/10.1016/j.ejsobi.2015.04.001>
- Antony-Babu, S., Murat, C., Deveau, A., Le Tacon, F., Frey-Klett, P., Uroz, S., 2013. An improved method compatible with metagenomic analyses to extract genomic DNA from soils in *Tuber melanosporum* orchards. *J. Appl. Microbiol.* 115, 163–170. <https://doi.org/10.1111/jam.12205>
- Aoyama, H., Saitoh, S., Fujii, S., Nagahama, H., Shinzato, N., Kaneko, N., Nakamori, T., 2015. A rapid method of non-destructive DNA extraction from individual springtails (Collembola). *Appl. Entomol. Zool.* 50, 419–425. <https://doi.org/10.1007/s13355-015-0340-0>
- Baini, F., Del Vecchio, M., Vizzari, L., Zapparoli, M., 2016. Can the efficiency of pitfall traps in collecting arthropods vary according to the used mixtures as bait? *Rendiconti Lincei* 27, 495–499. <https://doi.org/10.1007/s12210-016-0504-z>
- Bardgett, R.D., van der Putten, W.H., 2014. Belowground biodiversity and ecosystem functioning. *Nature* 515, 505–511. <https://doi.org/10.1038/nature13855>

- Barnes, M.A., Turner, C.R., 2016. The ecology of environmental DNA and implications for conservation genetics. *Conserv. Genet.* 17, 1–17. <https://doi.org/10.1007/s10592-015-0775-4>
- Battigelli, J.P., Spence, J.R., Langor, D.W., Berch, S.M., 2004. Short-term impact of forest soil compaction and organic matter removal on soil mesofauna density and oribatid mite diversity. *Can. J. For. Res.* 34, 1136–1149. <https://doi.org/10.1139/x03-267>
- Bedano, J.C., Domínguez, A., Arolfo, R., Wall, L.G., 2016. Effect of Good Agricultural Practices under no-till on litter and soil invertebrates in areas with different soil types. *Soil Tillage Res.* 158, 100–109. <https://doi.org/10.1016/j.still.2015.12.005>
- Behan-Pelletier, V.M., 1999. Oribatid mite biodiversity in agroecosystems: role for bioindication. *Agric. Ecosyst. Environ.* 74, 411–423.
- Bellinger, P.F., Christiansen, K.A., Janssens, F., 2018. Checklist of the Collembola of the World [WWW Document]. URL www.Collembola.org
- Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., Amend, A.S., Jumpponen, A., Unterseher, M., Kristiansson, E., Abarenkov, K., Bertrand, Y.J.K., Sanli, K., Eriksson, K.M., Vik, U., Veldre, V., Nilsson, R.H., 2013. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol. Evol.* 914–919. <https://doi.org/10.1111/2041-210X.12073>
- Berch, S.M., Battigelli, J.P., Hope, G.D., 2007. Responses of soil mesofauna communities and oribatid mite species to site preparation treatments in high-elevation cutblocks in southern British Columbia. *Pedobiologia* 51, 23–32. <https://doi.org/10.1016/j.pedobi.2006.12.001>
- Bird, S.B., Coulson, R.N., Fisher, R.F., 2004. Changes in soil and litter arthropod abundance following tree harvesting and site preparation in a loblolly pine (*Pinus taeda* L.) plantation. *For. Ecol. Manag.* 202, 195–208. <https://doi.org/10.1016/j.foreco.2004.07.023>
- Birkhofer, K., Dietrich, C., John, K., Schorpp, Q., Zaitsev, A.S., Wolters, V., 2016. Regional Conditions and Land-Use Alter the Potential Contribution of Soil Arthropods to Ecosystem Services in Grasslands. *Front. Ecol. Evol.* 3, 150. <https://doi.org/10.3389/fevo.2015.00150>
- Birky, C.W., 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu. Rev. Genet.* 35, 125–148. <https://doi.org/10.1146/annurev.genet.35.1.125>
- Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., Abebe, E., 2005. Defining operational taxonomic units using DNA barcode data. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 1935–1943. <https://doi.org/10.1098/rstb.2005.1725>

- Bokhorst, S., Kardol, P., Bellingham, P.J., Kooyman, R.M., Richardson, S.J., Schmidt, S., Wardle, D.A., 2017. Responses of communities of soil organisms and plants to soil aging at two contrasting long-term chronosequences. *Soil Biol. Biochem.* 106, 69–79. <https://doi.org/10.1016/j.soilbio.2016.12.014>
- Bradford, M.A., Tordoff, G.M., Eggers, T., Jones, T.H., Newington, J.E., 2002. Microbiota, fauna, and mesh size interactions in litter decomposition. *Oikos* 99, 317–323. <https://doi.org/10.1034/j.1600-0706.2002.990212.x>
- Braid, M.D., Daniels, L.M., Kitts, C.L., 2003. Removal of PCR inhibitors from soil DNA by chemical flocculation. *J. Microbiol. Methods* 52, 389–393. [https://doi.org/10.1016/S0167-7012\(02\)00070-2](https://doi.org/10.1016/S0167-7012(02)00070-2)
- Brescia, P., Banks, P., 2010. Micro-Volume PicoGreen® DNA Quantification using BioTek's Take3™ Multi-Volume Plate. BioTek Instruments, Inc., Winooski, VT.
- Buse, T., Ruess, L., Filser, J., 2014. Collembola gut passage shapes microbial communities in faecal pellets but not viability of dietary algal cells. *Chemoecology* 24, 79–84. <https://doi.org/10.1007/s00049-013-0145-y>
- Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* 11, 2639. <https://doi.org/10.1038/ismej.2017.119>
- Callejas-Chavero, A., Castaño-Meneses, G., Razo-González, M., Pérez-Velázquez, D., Palacios-Vargas, J.G., Flores-Martínez, A., 2015. Soil Microarthropods and Their Relationship to Higher Trophic Levels in the *Pedregal de San Angel* Ecological Reserve, Mexico. *J. Insect Sci.* 15, 59–59. <https://doi.org/10.1093/jisesa/iev039>
- Cannon, M.V., Hester, J., Shalkhauser, A., Chan, E.R., Logue, K., Small, S.T., Serre, D., 2016. *In silico* assessment of primers for eDNA studies using PrimerTree and application to characterize the biodiversity surrounding the Cuyahoga River. *Sci. Rep.* 6, 22908. <https://doi.org/10.1038/srep22908>
- Caruso, T., Pigino, G., Bernini, F., Bargagli, R., Migliorini, M., 2007. The Berger–Parker index as an effective tool for monitoring the biodiversity of disturbed soils: a case study on Mediterranean oribatid (Acari: Oribatida) assemblages. *Biodivers. Conserv.* 16, 3277–3285. <https://doi.org/10.1007/s10531-006-9137-3>
- Chauvat, M., Ponge, J.F., Wolters, V., 2007. Humus structure during a spruce forest rotation: quantitative changes and relationship to soil biota. *Eur. J. Soil Sci.* 58, 625–631. <https://doi.org/10.1111/j.1365-2389.2006.00847.x>
- Chen, H., Oram, N.J., Barry, K.E., Mommer, L., van Ruijven, J., de Kroon, H., Ebeling, A., Eisenhauer, N., Fischer, C., Gleixner, G., Gessler, A., González Macé, O., Hacker, N., Hildebrandt, A., Lange, M., Scherer-Lorenzen, M., Scheu, S., Oelmann, Y., Wagg, C., Wilcke, W., Wirth, C., Weigelt, A., 2017. Root chemistry and soil fauna, but not soil

- abiotic conditions explain the effects of plant diversity on root decomposition. *Oecologia* 185, 499–511. <https://doi.org/10.1007/s00442-017-3962-9>
- Cicconardi, F., Fanciulli, P.P., Emerson, B.C., 2013. Collembola, the biological species concept and the underestimation of global species richness. *Mol. Ecol.* 22, 5382–5396. <https://doi.org/10.1111/mec.12472>
- Cicconardi, F., Nardi, F., Emerson, B.C., Frati, F., Fanciulli, P.P., 2010. Deep phylogeographic divisions and long-term persistence of forest invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean basin. *Mol. Ecol.* 19, 386–400. <https://doi.org/10.1111/j.1365-294X.2009.04457.x>
- Coissac, E., Riaz, T., Puillandre, N., 2012. Bioinformatic challenges for DNA metabarcoding of plants and animals. *Mol. Ecol.* 21, 1834–1847. <https://doi.org/10.1111/j.1365-294X.2012.05550.x>
- Cole, L., Dromph, K.M., Boaglio, V., Bardgett, R.D., 2003. Effect of density and species richness of soil mesofauna on nutrient mineralisation and plant growth. *Biol. Fertil. Soils* 39, 337–343. <https://doi.org/10.1007/s00374-003-0702-6>
- Comtet, T., Sandionigi, A., Viard, F., Casiraghi, M., 2015. DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biol. Invasions* 17, 905–922. <https://doi.org/10.1007/s10530-015-0854-y>
- Coudrain, V., Hedde, M., Chauvat, M., Maron, P.-A., Bourgeois, E., Mary, B., Léonard, J., Ekelund, F., Villenave, C., Recous, S., 2016. Temporal differentiation of soil communities in response to arable crop management strategies. *Agric. Ecosyst. Environ.* 225, 12–21. <https://doi.org/10.1016/j.agee.2016.03.029>
- Cowart, D.A., Pinheiro, M., Mouchel, O., Maguer, M., Grall, J., Miné, J., Arnaud-Haond, S., 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. *PLoS One* 10, e0117562.
- Cristescu, M.E., 2014. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends Ecol. Evol.* 29, 566–571. <https://doi.org/10.1016/j.tree.2014.08.001>
- Crotty, F.V., Fychan, R., Scullion, J., Sanderson, R., Marley, C.L., 2015. Assessing the impact of agricultural forage crops on soil biodiversity and abundance. *Soil Biol. Biochem.* 91, 119–126. <https://doi.org/10.1016/j.soilbio.2015.08.036>
- Crowther, T.W., Stanton, D.W., Thomas, S.M., A’Bear, A.D., Hiscox, J., Jones, T.H., Voříšková, J., Baldrian, P., Boddy, L., 2013. Top-down control of soil fungal community composition by a globally distributed keystone consumer. *Ecology* 94, 2518–2528. <https://doi.org/10.1890/13-0197.1>

- Curry, C.J., Gibson, J.F., Shokralla, S., Hajibabaei, M., Baird, D.J., 2018. Identifying North American freshwater invertebrates using DNA barcodes: are existing COI sequence libraries fit for purpose? *Freshw. Sci.* 37, 178–189. <https://doi.org/10.1086/696613>
- D’Annibale, A., Sechi, V., Larsen, T., Christensen, S., Krogh, P.H., Eriksen, J., 2017. Does introduction of clover in an agricultural grassland affect the food base and functional diversity of Collembola? *Soil Biol. Biochem.* 112, 165–176. <https://doi.org/10.1016/j.soilbio.2017.05.010>
- De Deyn, G.B., Raaijmakers, C.E., Zoomer, H.R., Berg, M.P., de Ruiter, P.C., Verhoef, H.A., Bezemer, T.M., van der Putten, W.H., 2003. Soil invertebrate fauna enhances grassland succession and diversity. *Nature* 422, 711–713. <https://doi.org/10.1038/nature01548>
- Deagle, B.E., Jarman, S.N., Coissac, E., Pompanon, F., Taberlet, P., 2014. DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biol. Lett.* 10, 20140562. <https://doi.org/10.1098/rsbl.2014.0562>
- de la Peña, E., Baeten, L., Steel, H., Viaene, N., De Sutter, N., De Schrijver, A., Verheyen, K., 2016. Beyond plant-soil feedbacks: mechanisms driving plant community shifts due to land-use legacies in post-agricultural forests. *Funct. Ecol.* 30, 1073–1085. <https://doi.org/10.1111/1365-2435.12672>
- Decaëns, T., 2008. Priorities for conservation of soil animals. *CAB Rev. Perspect. Agric. Vet. Sci. Nutr. Nat. Resour.* 3, 14. <https://doi.org/10.1079/PAVSNNR20083014>
- Deharveng, L., 2004. Recent advances in Collembola systematics. *Pedobiologia* 48, 415–433. <https://doi.org/10.1016/j.pedobi.2004.08.001>
- Deiner, K., Bik, H.M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D.M., de Vere, N., Pfrender, M.E., Bernatchez, L., 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Mol. Ecol.* 26, 5872–5895. <https://doi.org/10.1111/mec.14350>
- Derycke, S., Vanaverbeke, J., Rigaux, A., Backeljau, T., Moens, T., 2010. Exploring the Use of Cytochrome Oxidase c Subunit I (COI) for DNA Barcoding of Free-Living Marine Nematodes. *PLoS ONE* 5, e13716. <https://doi.org/10.1371/journal.pone.0013716>
- deWaard, J.R., Landry, J.-F., Schmidt, B.C., Derhousoff, J., McLean, J.A., Humble, L.M., 2009. In the dark in a large urban park: DNA barcodes illuminate cryptic and introduced moth species. *Biodivers. Conserv.* 18, 3825–3839. <https://doi.org/10.1007/s10531-009-9682-7>
- Dirilgen, T., Juceviča, E., Melecis, V., Querner, P., Bolger, T., 2018. Analysis of spatial patterns informs community assembly and sampling requirements for Collembola in forest soils. *Acta Oecologica* 86, 23–30. <https://doi.org/10.1016/j.actao.2017.11.010>
- Donn, S., Griffiths, B.S., Neilson, R., Daniell, T.J., 2008. DNA extraction from soil nematodes for multi-sample community studies. *Appl. Soil Ecol.* 38, 20–26. <https://doi.org/10.1016/j.apsoil.2007.08.006>

- Drummond, A.J., Newcomb, R.D., Buckley, T.R., Xie, D., Dopheide, A., Potter, B.C., Heled, J., Ross, H.A., Tooman, L., Grosser, S., Park, D., Demetras, N.J., Stevens, M.I., Russell, J.C., Anderson, S.H., Carter, A., Nelson, N., 2015. Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience* 4, 46. <https://doi.org/10.1186/s13742-015-0086-1>
- Dutilleul, P., Stockwell, J.D., Frigon, D., Legendre, P., 2000. The Mantel Test versus Pearson's Correlation Analysis: Assessment of the Differences for Biological and Environmental Studies. *J. Agric. Biol. Environ. Stat.* 5, 131. <https://doi.org/10.2307/1400528>
- Edgar, R.C., 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv*. <https://doi.org/10.1101/081257>
- Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Elbrecht, V., Leese, F., 2015. Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass—Sequence Relationships with an Innovative Metabarcoding Protocol. *PLOS ONE* 10, e0130324. <https://doi.org/10.1371/journal.pone.0130324>
- Elbrecht, V., Peinert, B., Leese, F., 2017. Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecol. Evol.* 7, 6918–6926. <https://doi.org/10.1002/ece3.3192>
- Emilson, C.E., Thompson, D.G., Venier, L.A., Porter, T.M., Swystun, T., Chartrand, D., Capell, S., Hajibabaei, M., 2017. DNA metabarcoding and morphological macroinvertebrate metrics reveal the same changes in boreal watersheds across an environmental gradient. *Sci. Rep.* 7. <https://doi.org/10.1038/s41598-017-13157-x>
- Farská, J., Prejzková, K., Rusek, J., 2014. Management intensity affects traits of soil microarthropod community in montane spruce forest. *Appl. Soil Ecol.* 75, 71–79. <https://doi.org/10.1016/j.apsoil.2013.11.003>
- Ficetola, G.F., Miaud, C., Pompanon, F., Taberlet, P., 2008. Species detection using environmental DNA from water samples. *Biol. Lett.* 4, 423–425. <https://doi.org/10.1098/rsbl.2008.0118>
- Ficetola, G.F., Pansu, J., Bonin, A., Coissac, E., Giguët-Covex, C., De Barba, M., Gielly, L., Lopes, C.M., Boyer, F., Pompanon, F., Rayé, G., Taberlet, P., 2015. Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Mol. Ecol. Resour.* 15, 543–556. <https://doi.org/10.1111/1755-0998.12338>
- Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R.A., Felts, B., Rayhawk, S., Knight, R., Rohwer, F., Jackson, R.B., 2007. Metagenomic and Small-Subunit rRNA Analyses Reveal the Genetic Diversity of Bacteria, Archaea, Fungi, and Viruses in Soil. *Appl. Environ. Microbiol.* 73, 7059–7066. <https://doi.org/10.1128/AEM.00358-07>

- Fierer, N., Strickland, M.S., Liptzin, D., Bradford, M.A., Cleveland, C.C., 2009. Global patterns in belowground communities. *Ecol. Lett.* 12, 1238–1249. <https://doi.org/10.1111/j.1461-0248.2009.01360.x>
- Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* 3, 294–299.
- Friberg, H., Lagerlöf, J., Rämert, B., 2005. Influence of soil fauna on fungal plant pathogens in agricultural and horticultural systems. *Biocontrol Sci. Technol.* 15, 641–658. <https://doi.org/10.1080/09583150500086979>
- Galtier, N., Nabholz, B., GléMin, S., Hurst, G.D.D., 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* 18, 4541–4550. <https://doi.org/10.1111/j.1365-294X.2009.04380.x>
- Geisen, S., Laros, I., Vizcaíno, A., Bonkowski, M., de Groot, G.A., 2015. Not all are free-living: high-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil metazoa. *Mol. Ecol.* 24, 4556–4569. <https://doi.org/10.1111/mec.13238>
- Geisen, S., Mitchell, E.A.D., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., Fernández, L.D., Jousset, A., Krashevskaya, V., Singer, D., Spiegel, F.W., Walochnik, J., Lara, E., 2018. Soil protists: a fertile frontier in soil biology research. *FEMS Microbiol. Rev.* 42, 293–323. <https://doi.org/10.1093/femsre/fuy006>
- Gibson, J., Shokralla, S., Porter, T.M., King, I., van Konynenburg, S., Janzen, D.H., Hallwachs, W., Hajibabaei, M., 2014. Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasytematics. *Proc. Natl. Acad. Sci.* 111, 8007–8012. <https://doi.org/10.1073/pnas.1406468111>
- Gibson, J.F., Shokralla, S., Curry, C., Baird, D.J., Monk, W.A., King, I., Hajibabaei, M., 2015. Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. *PloS One* 10, e0138432.
- Goslee, S., Urban, D., 2017. *ecodist: Dissimilarity-Based Functions for Ecological Analysis*.
- Gulvik, M., 2007. Mites (Acari) as indicators of soil biodiversity and land use monitoring: a review. *Pol. J. Ecol.* 55, 415–440.
- Hajibabaei, M., deWaard, J.R., Ivanova, N.V., Ratnasingham, S., Dooh, R.T., Kirk, S.L., Mackie, P.M., Hebert, P.D.N., 2005. Critical factors for assembling a high volume of DNA barcodes. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 1959–1967. <https://doi.org/10.1098/rstb.2005.1727>
- Hajibabaei, M., Spall, J.L., Shokralla, S., van Konynenburg, S., 2012. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. *BMC Ecol.* 12, 28. <https://doi.org/10.1186/1472-6785-12-28>

- Hamilton, H.C., Strickland, M.S., Wickings, K., Bradford, M.A., Fierer, N., 2009. Surveying soil faunal communities using a direct molecular approach. *Soil Biol. Biochem.* 41, 1311–1314. <https://doi.org/10.1016/j.soilbio.2009.03.021>
- Harrop-Archibald, H., Didham, R.K., Standish, R.J., Tibbett, M., Hobbs, R.J., 2016. Mechanisms linking fungal conditioning of leaf litter to detritivore feeding activity. *Soil Biol. Biochem.* 93, 119–130. <https://doi.org/10.1016/j.soilbio.2015.10.021>
- Hebert, P.D.N., Cywinska, A., Ball, S.L., others, 2003a. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B Biol. Sci.* 270, 313–321.
- Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14812–14817. <https://doi.org/10.1073/pnas.0406166101>
- Hebert, P.D.N., Ratnasingham, S., de Waard, J.R., 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. B Biol. Sci.* 270, S96–S99. <https://doi.org/10.1098/rsbl.2003.0025>
- Heethoff, M., Laumann, M., Weigmann, G., Rasputnig, G., 2011. Integrative taxonomy: combining morphological, molecular and chemical data for species delineation in the parthenogenetic *Trhypochthonius tectorum* complex (Acari, Oribatida, Trhypochthoniidae). *Front. Zool.* 8, 2.
- Heneghan, L., Bolger, T., 1998. Soil microarthropod contribution to forest ecosystem processes: the importance of observational scale. *Plant Soil* 205, 113–124.
- Hernández-Triana, L.M., Prosser, S.W., Rodríguez-Perez, M.A., Chaverri, L.G., Hebert, P.D.N., Ryan Gregory, T., 2014. Recovery of DNA barcodes from blackfly museum specimens (Diptera: Simuliidae) using primer sets that target a variety of sequence lengths. *Mol. Ecol. Resour.* 14, 508–518. <https://doi.org/10.1111/1755-0998.12208>
- Hishi, T., Takeda, H., 2008. Soil microarthropods alter the growth and morphology of fungi and fine roots of *Chamaecyparis obtusa*. *Pedobiologia* 52, 97–110. <https://doi.org/10.1016/j.pedobi.2008.04.003>
- Hogg, I.D., Hebert, P.D., 2004. Biological identification of springtails (Hexapoda: Collembola) from the Canadian Arctic, using mitochondrial DNA barcodes. *Can. J. Zool.* 82, 749–754. <https://doi.org/10.1139/z04-041>
- Hopkin, S.P., 1997. *Biology of the Springtails (Insecta: Collembola)*. Oxford University Press, United Kingdom.
- Ingimarsdóttir, M., Caruso, T., Ripa, J., Magnúsdóttir, Ó.B., Migliorini, M., Hedlund, K., 2012. Primary assembly of soil communities: disentangling the effect of dispersal and local environment. *Oecologia* 170, 745–754. <https://doi.org/10.1007/s00442-012-2334-8>

- Ji, Y., Ashton, L., Pedley, S.M., Edwards, D.P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P.M., Woodcock, P., Edwards, F.A., Larsen, T.H., Hsu, W.W., Benedick, S., Hamer, K.C., Wilcove, D.S., Bruce, C., Wang, X., Levi, T., Lott, M., Emerson, B.C., Yu, D.W., 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecol. Lett.* 16, 1245–1257. <https://doi.org/10.1111/ele.12162>
- Jinbo, U., Kato, T., Ito, M., 2011. Current progress in DNA barcoding and future implications for entomology: DNA barcoding for entomology. *Entomol. Sci.* 14, 107–124. <https://doi.org/10.1111/j.1479-8298.2011.00449.x>
- Karimi, B., Maron, P.A., Chemidlin-Prevost Boure, N., Bernard, N., Gilbert, D., Ranjard, L., 2017. Microbial diversity and ecological networks as indicators of environmental quality. *Environ. Chem. Lett.* 15, 265–281. <https://doi.org/10.1007/s10311-017-0614-6>
- King, R.A., Tibble, A.L., Symondson, W.O.C., 2008. Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. *Mol. Ecol.* 17, 4684–4698. <https://doi.org/10.1111/j.1365-294X.2008.03931.x>
- Kirilenko, A.P., Sedjo, R.A., 2007. Climate change impacts on forestry. *Proc. Natl. Acad. Sci.* 104, 19697–19702. <https://doi.org/10.1073/pnas.0701424104>
- Knowlton, N., Weigt, L.A., 1998. New dates and new rates for divergence across the Isthmus of Panama. *Proc. R. Soc. Lond. B Biol. Sci.* 265, 2257–2263. <https://doi.org/10.1098/rspb.1998.0568>
- Kõljag, U., Nilsson, R.H., Abarenkov, K., Tedersoo, L., Taylor, A.F.S., Bahram, M., Bates, S.T., Bruns, T.D., Bengtsson-Palme, J., Callaghan, T.M., Douglas, B., Drenkhan, T., Eberhardt, U., Dueñas, M., Grebenc, T., Griffith, G.W., Hartmann, M., Kirk, P.M., Kohout, P., Larsson, E., Lindahl, B.D., Lücking, R., Martín, M.P., Matheny, P.B., Nguyen, N.H., Niskanen, T., Oja, J., Peay, K.G., Peintner, U., Peterson, M., Põldmaa, K., Saag, L., Saar, I., Schüßler, A., Scott, J.A., Senés, C., Smith, M.E., Suija, A., Taylor, D.L., Telleria, M.T., Weiss, M., Larsson, K.-H., 2013. Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* 22, 5271–5277. <https://doi.org/10.1111/mec.12481>
- Krogh, P.H., Miljøundersøgelser, D., 2009. Toxicity testing with the collembolans *Folsomia fimetaria* and *Folsomia candida* and the results of a ringtest. Danish Environmental Protection Agency.
- Krsek, M., Wellington, E.M.H., 1999. Comparison of different methods for the isolation and purification of total community DNA from soil. *J. Microbiol. Methods* 39, 1–16. [https://doi.org/10.1016/S0167-7012\(99\)00093-7](https://doi.org/10.1016/S0167-7012(99)00093-7)
- Kwiaton, M., Hazlett, P., Morris, D., Fleming, R., Webster, K., Venier, L., Aubin, I., 2014. Island Lake Biomass Harvest Research and Demonstration Area: Establishment Report (No. GLC-X-11). Natural Resources Canada, Canadian Forestry Service, Great Lakes Forestry Centre.

- Laumann, M., Norton, R.A., Weigmann, G., Scheu, S., Maraun, M., Heethoff, M., 2007. Speciation in the parthenogenetic oribatid mite genus *Tectocepheus* (Acari, Oribatida) as indicated by molecular phylogeny. *Pedobiologia* 51, 111–122. <https://doi.org/10.1016/j.pedobi.2007.02.001>
- Legendre, P., Fortin, M.J., 1989. Spatial pattern and ecological analysis. *Vegetatio* 80, 107–138.
- Lehmitz, R., Decker, P., 2017. The nuclear 28S gene fragment D3 as species marker in oribatid mites (Acari, Oribatida) from German peatlands. *Exp. Appl. Acarol.* 71, 259–276. <https://doi.org/10.1007/s10493-017-0126-x>
- Lilleskov, E.A., Bruns, T.D., 2005. Spore dispersal of a resupinate ectomycorrhizal fungus, *Tomentella subulilacina*, via soil food webs. *Mycologia* 97, 762–769. <https://doi.org/10.1080/15572536.2006.11832767>
- Lindo, Z., Gonzalez, A., 2010. The Bryosphere: An Integral and Influential Component of the Earth's Biosphere. *Ecosystems* 13, 612–627. <https://doi.org/10.1007/s10021-010-9336-3>
- Lindo, Z., N. Winchester, N., 2008. Scale dependent diversity patterns in arboreal and terrestrial oribatid mite (Acari: Oribatida) communities. *Ecography* 31, 53–60. <https://doi.org/10.1111/j.2007.0906-7590.05320.x>
- Lindo, Z., Visser, S., 2004. Forest floor microarthropod abundance and oribatid mite (Acari: Oribatida) composition following partial and clear-cut harvesting in the mixedwood boreal forest. *Can. J. For. Res.* 34, 998–1006. <https://doi.org/10.1139/x03-284>
- Lindo, Z., Whiteley, J., Gonzalez, A., 2012. Traits explain community disassembly and trophic contraction following experimental environmental change. *Glob. Change Biol.* 18, 2448–2457. <https://doi.org/10.1111/j.1365-2486.2012.02725.x>
- Maaß, S., Caruso, T., Rillig, M.C., 2015. Functional role of microarthropods in soil aggregation. *Pedobiologia* 58, 59–63. <https://doi.org/10.1016/j.pedobi.2015.03.001>
- Macher, J.-N., Macher, T.-H., Leese, F., 2017. Combining NCBI and BOLD databases for OTU assignment in metabarcoding and metagenomic datasets: The BOLD_NCBI_Merger. *Metabarcoding Metagenomics* 1, e22262. <https://doi.org/10.3897/mbmg.1.22262>
- Malcicka, M., Berg, M.P., Ellers, J., 2017. Ecomorphological adaptations in Collembola in relation to feeding strategies and microhabitat. *Eur. J. Soil Biol.* 78, 82–91. <https://doi.org/10.1016/j.ejsobi.2016.12.004>
- Malmström, A., 2012. Life-history traits predict recovery patterns in Collembola species after fire: A 10 year study. *Appl. Soil Ecol.* 56, 35–42. <https://doi.org/10.1016/j.apsoil.2012.02.007>
- Malmström, A., Persson, T., Ahlström, K., Gongalsky, K.B., Bengtsson, J., 2009. Dynamics of soil meso- and macrofauna during a 5-year period after clear-cut burning in a boreal forest. *Appl. Soil Ecol.* 43, 61–74. <https://doi.org/10.1016/j.apsoil.2009.06.002>

- Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27, 209–220.
- Maraun, M., Heethoff, M., Schneider, K., Scheu, S., Weigmann, G., Cianciolo, J., Thomas, R.H., Norton, R.A., 2004. Molecular phylogeny of oribatid mites (Oribatida, Acari): evidence for multiple radiations of parthenogenetic lineages. *Exp. Appl. Acarol.* 33, 183–201. <https://doi.org/10.1023/B:APPA.0000032956.60108.6d>
- Maraun, M., Martens, H., Migge, S., Theenhaus, A., Scheu, S., 2003a. Adding to “the enigma of soil animal diversity”: fungal feeders and saprophagous soil invertebrates prefer similar food substrates. *Eur. J. Soil Biol.* 39, 85–95. [https://doi.org/10.1016/S1164-5563\(03\)00006-2](https://doi.org/10.1016/S1164-5563(03)00006-2)
- Maraun, M., Salamon, J.-A., Schneider, K., Schaefer, M., Scheu, S., 2003b. Oribatid mite and collembolan diversity, density and community structure in a moder beech forest (*Fagus sylvatica*): effects of mechanical perturbations. *Soil Biol. Biochem.* 35, 1387–1394. [https://doi.org/10.1016/S0038-0717\(03\)00218-9](https://doi.org/10.1016/S0038-0717(03)00218-9)
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. <https://doi.org/10.14806/ej.17.1.200>
- Martinsson, S., Rota, E., Erséus, C., 2015. Revision of *Cognettia* (Clitellata, Enchytraeidae): re-establishment of *Chamaedrillus* and description of cryptic species in the *sphagnetorum* complex. *Syst. Biodivers.* 13, 257–277. <https://doi.org/10.1080/14772000.2014.986555>
- McKee, A.M., Spear, S.F., Pierson, T.W., 2015. The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biol. Conserv.* 183, 70–76. <https://doi.org/10.1016/j.biocon.2014.11.031>
- McMurdie, P.J., Holmes, S., 2017. phyloseq: Handling and analysis of high-throughput microbiome census data.
- McSorley, R., Walter, D.E., 1991. Comparison of soil extraction methods for nematodes and microarthropods. *Agric. Ecosyst. Environ.* 34, 201–207. [https://doi.org/10.1016/0167-8809\(91\)90350-5](https://doi.org/10.1016/0167-8809(91)90350-5)
- Menkis, A., Burokienė, D., Gaitnieks, T., Uotila, A., Johannesson, H., Rosling, A., Finlay, R.D., Stenlid, J., Vasaitis, R., 2012. Occurrence and impact of the root-rot biocontrol agent *Phlebiopsis gigantea* on soil fungal communities in *Picea abies* forests of northern Europe. *FEMS Microbiol. Ecol.* 81, 438–445. <https://doi.org/10.1111/j.1574-6941.2012.01366.x>
- Miller, D.N., Bryant, J.E., Madsen, E.L., Ghiorse, W.C., 1999. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl. Environ. Microbiol.* 65, 4715–4724. <https://doi.org/10.1128/AEM.65.12.4715-4724.1999>

- Minor, M.A., 2011. Spatial patterns and local diversity in soil oribatid mites (Acari: Oribatida) in three pine plantation forests. *Eur. J. Soil Biol.* 47, 122–128.
<https://doi.org/10.1016/j.ejsobi.2011.01.003>
- Morlon, H., Chuyong, G., Condit, R., Hubbell, S., Kenfack, D., Thomas, D., Valencia, R., Green, J.L., 2008. A general framework for the distance-decay of similarity in ecological communities. *Ecol. Lett.* 11, 904–917. <https://doi.org/10.1111/j.1461-0248.2008.01202.x>
- Oelbermann, K., Scheu, S., 2010. Trophic guilds of generalist feeders in soil animal communities as indicated by stable isotope analysis ($^{15}\text{N}/^{14}\text{N}$). *Bull. Entomol. Res.* 100, 511–520.
<https://doi.org/10.1017/S0007485309990587>
- Ojala, R., Huhta, V., 2001. Dispersal of microarthropods in forest soil. *Pedobiologia* 45, 443–450. [https://doi.org/0031-4056/01/45/05-443\\$15.00/0](https://doi.org/0031-4056/01/45/05-443$15.00/0)
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2013. *vegan: Community Ecology Package*.
- Oliverio, A.M., Gan, H., Wickings, K., Fierer, N., 2018. A DNA metabarcoding approach to characterize soil arthropod communities. *Soil Biol. Biochem.* 125, 37–43.
<https://doi.org/10.1016/j.soilbio.2018.06.026>
- Op De Beeck, M., Lievens, B., Busschaert, P., Declerck, S., Vangronsveld, J., Colpaert, J.V., 2014. Comparison and Validation of Some ITS Primer Pairs Useful for Fungal Metabarcoding Studies. *PLoS ONE* 9, e97629.
<https://doi.org/10.1371/journal.pone.0097629>
- Orgiazzi, A., Dunbar, M.B., Panagos, P., de Groot, G.A., Lemanceau, P., 2015. Soil biodiversity and DNA barcodes: opportunities and challenges. *Soil Biol. Biochem.* 80, 244–250.
<https://doi.org/10.1016/j.soilbio.2014.10.014>
- Parisi, V., Menta, C., Gardi, C., Jacomini, C., Mozzanica, E., 2005. Microarthropod communities as a tool to assess soil quality and biodiversity: a new approach in Italy. *Agric. Ecosyst. Environ.* 105, 323–333. <https://doi.org/10.1016/j.agee.2004.02.002>
- Perdomo, G., Evans, A., Maraun, M., Sunnucks, P., Thompson, R., 2012. Mouthpart morphology and trophic position of microarthropods from soils and mosses are strongly correlated. *Soil Biol. Biochem.* 53, 56–63. <https://doi.org/10.1016/j.soilbio.2012.05.002>
- Polz, M.F., Cavanaugh, C.M., 1998. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* 64, 3724–3730.
- Ponge, J.F., Salmon, S., 2013. Spatial and taxonomic correlates of species and species trait assemblages in soil invertebrate communities. *Pedobiologia* 56, 129–136.
<https://doi.org/10.1016/j.pedobi.2013.02.001>

- Porco, D., Bedos, A., Greenslade, P., Janion, C., Skarżyński, D., Stevens, M.I., Jansen van Vuuren, B., Deharveng, L., 2012. Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. *Invertebr. Syst.* 26, 470. <https://doi.org/10.1071/IS12026>
- Porco, D., Decaëns, T., Deharveng, L., James, S.W., Skarżyński, D., Erséus, C., Butt, K.R., Richard, B., Hebert, P.D.N., 2013. Biological invasions in soil: DNA barcoding as a monitoring tool in a multiple taxa survey targeting European earthworms and springtails in North America. *Biol. Invasions* 15, 899–910. <https://doi.org/10.1007/s10530-012-0338-2>
- Porco, D., Rougerie, R., Deharveng, L., Hebert, P., 2010. Coupling non-destructive DNA extraction and voucher retrieval for small soft-bodied Arthropods in a high-throughput context: the example of Collembola. *Mol. Ecol. Resour.* 10, 942–945. <https://doi.org/10.1111/j.1755-0998.2010.02839.x>
- Porco, D., Skarżyński, D., Decaëns, T., Hebert, P.D.N., Deharveng, L., 2014. Barcoding the Collembola of Churchill: a molecular taxonomic reassessment of species diversity in a sub-Arctic area. *Mol. Ecol. Resour.* 14, 249–261. <https://doi.org/10.1111/1755-0998.12172>
- Porter, T.M., Hajibabaei, M., 2018a. Over 2.5 million COI sequences in GenBank and growing. *bioRxiv* 353904. <https://doi.org/10.1101/353904>
- Porter, T.M., Hajibabaei, M., 2018b. Automated high throughput animal CO1 metabarcode classification. *Sci. Rep.* 8. <https://doi.org/10.1038/s41598-018-22505-4>
- Potapov, A.A., Semenina, E.E., Korotkevich, A.Y., Kuznetsova, N.A., Tiunov, A.V., 2016. Connecting taxonomy and ecology: Trophic niches of collembolans as related to taxonomic identity and life forms. *Soil Biol. Biochem.* 101, 20–31. <https://doi.org/10.1016/j.soilbio.2016.07.002>
- Potapov, A.M., Tiunov, A.V., 2016. Stable isotope composition of mycophagous collembolans versus mycotrophic plants: Do soil invertebrates feed on mycorrhizal fungi? *Soil Biol. Biochem.* 93, 115–118. <https://doi.org/10.1016/j.soilbio.2015.11.001>
- Prasifka, J.R., Lopez, M.D., Hellmich, R.L., Lewis, L.C., Dively, G.P., 2007. Comparison of pitfall traps and litter bags for sampling ground-dwelling arthropods: Comparison of pitfall trap and litter bag sampling. *J. Appl. Entomol.* 131, 115–120. <https://doi.org/10.1111/j.1439-0418.2006.01141.x>
- Princz, J.I., Behan-Pelletier, V.M., Scroggins, R.P., Siciliano, S.D., 2010. Oribatid mites in soil toxicity testing-the use of *Oppia nitens* (C.L. Koch) as a new test species. *Environ. Toxicol. Chem.* 29, 971–979. <https://doi.org/10.1002/etc.98>
- R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

- Ratnasingham, S., Hebert, P.D., 2007. BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Mol. Ecol. Resour.* 7, 355–364. <https://doi.org/10.1111/j.1471-8286.2006.01678.x>
- Ratnasingham, S., Hebert, P.D.N., 2013. A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. *PLoS ONE* 8, e66213. <https://doi.org/10.1371/journal.pone.0066213>
- Renker, C., Otto, P., Schneider, K., Zimdars, B., Maraun, M., Buscot, F., 2005. Oribatid Mites as Potential Vectors for Soil Microfungi: Study of Mite-Associated Fungal Species. *Microb. Ecol.* 50, 518–528. <https://doi.org/10.1007/s00248-005-5017-8>
- Resch, M.C., Shrubovych, J., Bartel, D., Szucsich, N.U., Timelthaler, G., Bu, Y., Walzl, M., Pass, G., 2014. Where Taxonomy Based on Subtle Morphological Differences Is Perfectly Mirrored by Huge Genetic Distances: DNA Barcoding in Protura (Hexapoda). *PLoS ONE* 9, e90653. <https://doi.org/10.1371/journal.pone.0090653>
- Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G., Triplett, E.W., 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* 1, 283–290. <https://doi.org/10.1038/ismej.2007.53>
- Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584. <https://doi.org/10.7717/peerj.2584>
- Roh, C., Villatte, F., Kim, B.-G., Schmid, R.D., 2006. Comparative study of methods for extraction and purification of environmental DNA from soil and sludge samples. *Appl. Biochem. Biotechnol.* 134, 97–112. <https://doi.org/10.1007/s12010-006-0097-3>
- Rougerie, R., Decaëns, T., Deharveng, L., Porco, D., James, S.W., Chang, C.-H., Richard, B., Potapov, M., Suhardjono, Y., Hebert, P.D., 2009. DNA barcodes for soil animal taxonomy. *Pesqui. Agropecu. Bras.* 44, 789–802. <https://doi.org/10.1590/S0100-204X2009000800002>
- Rousseau, L., 2018. Réponses des communautés de mésofaune du sol forestier à la récolte intensive de la biomasse ligneuse dans des peuplements de pin gris du nord-est de l'Ontario. (Thèse de doctorat en Sciences biologiques). Université du Québec à Montréal, Montréal.
- Rousseau, L., Venier, L., Hazlett, P., Fleming, R., Morris, D., Handa, I.T., 2018. Forest floor mesofauna communities respond to a gradient of biomass removal and soil disturbance in a boreal jack pine (*Pinus banksiana*) stand of northeastern Ontario (Canada). *For. Ecol. Manag.* 407, 155–165. <https://doi.org/10.1016/j.foreco.2017.08.054>
- Rusek, J., 1998. Biodiversity of Collembola and their functional role in the ecosystem. *Biodivers. Conserv.* 7, 1207–1219. <https://doi.org/10.1023/A:1008887817883>

- Sackett, T.E., Classen, A.T., Sanders, N.J., 2010. Linking soil food web structure to above- and belowground ecosystem processes: a meta-analysis. *Oikos* 119, 1984–1992. <https://doi.org/10.1111/j.1600-0706.2010.18728.x>
- Sagova-Mareckova, M., Cermak, L., Novotna, J., Plhachova, K., Forstova, J., Kopecky, J., 2008. Innovative Methods for Soil DNA Purification Tested in Soils with Widely Differing Characteristics. *Appl. Environ. Microbiol.* 74, 2902–2907. <https://doi.org/10.1128/AEM.02161-07>
- Schäffer, S., Pfingstl, T., Koblmüller, S., Winkler, K.A., Sturmbauer, C., Krisper, G., 2010. Phylogenetic analysis of European *Scutovertex* mites (Acari, Oribatida, Scutoverticidae) reveals paraphyly and cryptic diversity - A molecular genetic and morphological approach. *Mol. Phylogenet. Evol.* 55, 677–688. <https://doi.org/10.1016/j.ympev.2009.11.025>
- Schmidt, P.-A., Bálint, M., Greshake, B., Bandow, C., Römbke, J., Schmitt, I., 2013. Illumina metabarcoding of a soil fungal community. *Soil Biol. Biochem.* 65, 128–132. <https://doi.org/10.1016/j.soilbio.2013.05.014>
- Schneider, K., Migge, S., Norton, R.A., Scheu, S., Langel, R., Reineking, A., Maraun, M., 2004. Trophic niche differentiation in soil microarthropods (Oribatida, Acari): evidence from stable isotope ratios ($^{15}\text{N}/^{14}\text{N}$). *Soil Biol. Biochem.* 36, 1769–1774. <https://doi.org/10.1016/j.soilbio.2004.04.033>
- Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Bolchacova, E., Voigt, K., Crous, P.W., others, 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci.* 109, 6241–6246. <https://doi.org/10.1073/pnas.1117018109>
- Sechi, V., De Goede, R.G.M., Rutgers, M., Brussaard, L., Mulder, C., 2017. A community trait-based approach to ecosystem functioning in soil. *Agric. Ecosyst. Environ.* 239, 265–273. <https://doi.org/10.1016/j.agee.2017.01.036>
- Shaw, P., Benefer, C.M., 2015. Development of a barcoding database for the UK Collembola: early results. *Soil Org.* 87, 197–202.
- Shaw, P., Faria, C., Emerson, B., 2013. Updating taxonomic biogeography in the light of new methods – examples from Collembola. *Soil Org.* 85, 161–170.
- Siddiky, M.R.K., Schaller, J., Caruso, T., Rillig, M.C., 2012. Arbuscular mycorrhizal fungi and collembola non-additively increase soil aggregation. *Soil Biol. Biochem.* 47, 93–99. <https://doi.org/10.1016/j.soilbio.2011.12.022>
- Siira-Pietikäinen, A., Haimi, J., 2009. Changes in soil fauna 10 years after forest harvestings: Comparison between clear felling and green-tree retention methods. *For. Ecol. Manag.* 258, 332–338. <https://doi.org/10.1016/j.foreco.2009.04.024>

- Siira-Pietikäinen, A., Pietikäinen, J., Fritze, H., Haimi, J., 2001. Short-term responses of soil decomposer communities to forest management: clear felling versus alternative forest harvesting methods. *Can. J. For. Res.* 31, 88–99. <https://doi.org/10.1139-cjfr-31-1-88>
- Simon, C., Daniel, R., 2011. Metagenomic Analyses: Past and Future Trends. *Appl. Environ. Microbiol.* 77, 1153–1161. <https://doi.org/10.1128/AEM.02345-10>
- Smith, M.A., Fisher, B.L., 2009. Invasions, DNA barcodes, and rapid biodiversity assessment using ants of Mauritius. *Front. Zool.* 6, 31. <https://doi.org/10.1186/1742-9994-6-31>
- Socarrás, A., 2013. Soil mesofauna: biological indicator of soil quality. *Pastos Forrajes* 36, 14–21.
- Soil Classification Working Group, 1998. The Canadian System of Soil Classification, 3rd ed. NRC Research Press, Canada.
- Soininen, J., McDonald, R., Hillebrand, H., 2007. The distance decay of similarity in ecological communities. *Ecography* 30, 3–12. <https://doi.org/10.1111/j.2006.0906-7590.04817.x>
- Soler, R., Van der Putten, W.H., Harvey, J.A., Vet, L.E.M., Dicke, M., Bezemer, T.M., 2012. Root Herbivore Effects on Aboveground Multitrophic Interactions: Patterns, Processes and Mechanisms. *J. Chem. Ecol.* 38, 755–767. <https://doi.org/10.1007/s10886-012-0104-z>
- Soong, J.L., Vandegehuchte, M.L., Horton, A.J., Nielsen, U.N., Denef, K., Shaw, E.A., de Tomasel, C.M., Parton, W., Wall, D.H., Cotrufo, M.F., 2016. Soil microarthropods support ecosystem productivity and soil C accrual: Evidence from a litter decomposition study in the tallgrass prairie. *Soil Biol. Biochem.* 92, 230–238. <https://doi.org/10.1016/j.soilbio.2015.10.014>
- Staadén, S., Milcu, A., Rohlfs, M., Scheu, S., 2011. Olfactory cues associated with fungal grazing intensity and secondary metabolite pathway modulate Collembola foraging behaviour. *Soil Biol. Biochem.* 43, 1411–1416. <https://doi.org/10.1016/j.soilbio.2010.10.002>
- Steinaker, D.F., Wilson, S.D., 2008. Scale and density dependent relationships among roots, mycorrhizal fungi and collembola in grassland and forest. *Oikos* 117, 703–710. <https://doi.org/10.1111/j.2008.0030-1299.16452.x>
- Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., Breiner, H.-W., Richards, T.A., 2010. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* 19, 21–31. <https://doi.org/10.1111/j.1365-294X.2009.04480.x>
- Sun, X., Zhang, F., Ding, Y., Davies, T.W., Li, Y., Wu, D., 2017. Delimiting species of *Protaphorura* (Collembola: Onychiuridae): integrative evidence based on morphology, DNA sequences and geography. *Sci. Rep.* 7. <https://doi.org/10.1038/s41598-017-08381-4>

- Suzuki, M.T., Giovannoni, S.J., 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl. Environ. Microbiol.* 62, 625–630.
- Swenson, N.G., Jones, F.A., 2017. Community transcriptomics, genomics and the problem of species co-occurrence. *J. Ecol.* 105, 563–568. <https://doi.org/10.1111/1365-2745.12771>
- Taberlet, P., Coissac, E., Hajibabaei, M., Rieseberg, L.H., 2012a. Environmental DNA. *Mol. Ecol.* 21, 1789–1793. <https://doi.org/10.1111/j.1365-294X.2012.05542.x>
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willerslev, E., 2012b. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* 21, 2045–2050. <https://doi.org/10.1111/j.1365-294X.2012.05470.x>
- Tang, C.Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T.G., Fontaneto, D., 2012. The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proc. Natl. Acad. Sci.* 109, 16208–16212. <https://doi.org/10.1073/pnas.1209160109>
- Técher, D., Martinez-Chois, C., D’Innocenzo, M., Laval-Gilly, P., Bennasroune, A., Foucaud, L., Falla, J., 2010. Novel perspectives to purify genomic DNA from high humic acid content and contaminated soils. *Sep. Purif. Technol.* 75, 81–86. <https://doi.org/10.1016/j.seppur.2010.07.014>
- Tedersoo, L., Anslan, S., Bahram, M., Põlme, S., Riit, T., Liiv, I., Kõljalg, U., Kisand, V., Nilsson, H., Hildebrand, F., Bork, P., Abarenkov, K., 2015. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys* 10, 1–43. <https://doi.org/10.3897/mycokeys.10.4852>
- Tedersoo, L., Tooming-Klunderud, A., Anslan, S., 2018. PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytol.* 217, 1370–1385. <https://doi.org/10.1111/nph.14776>
- Terrat, S., Plassart, P., Bourgeois, E., Ferreira, S., Dequiedt, S., Adele-Dit-De-Renseville, N., Lemanceau, P., Bispo, A., Chabbi, A., Maron, P.-A., Ranjard, L., 2015. Meta-barcoded evaluation of the ISO standard 11063 DNA extraction procedure to characterize soil bacterial and fungal community diversity and composition: Meta-barcoded evaluation of the ISO-11063 standard. *Microb. Biotechnol.* 8, 131–142. <https://doi.org/10.1111/1751-7915.12162>
- Thiffault, E., Paré, D., Brais, S., Titus, B.D., 2010. Intensive biomass removals and site productivity in Canada: a review of relevant issues. *For. Chron.* 86, 36–42. <https://doi.org/10.5558/tfc86036-1>
- Treonis, A.M., Unangst, S.K., Kepler, R.M., Buyer, J.S., Cavigelli, M.A., Mirsky, S.B., Maul, J.E., 2018. Characterization of soil nematode communities in three cropping systems through morphological and DNA metabarcoding approaches. *Sci. Rep.* 8. <https://doi.org/10.1038/s41598-018-20366-5>

- Walter, D.E., Proctor, H.C., 2013. Mites: Ecology, Evolution and Behaviour, 2nd ed. Springer Science & Business Media.
- Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. <https://doi.org/10.1128/AEM.00062-07>
- White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics, in: *PCR Protocols: A Guide to Methods and Applications*. Academic Press, New York, USA, pp. 315–322.
- Wickham, H., Chang, W., 2016. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.
- Wickings, K., Grandy, A.S., 2011. The oribatid mite *Scheloribates moestus* (Acari: Oribatida) alters litter chemistry and nutrient cycling during decomposition. *Soil Biol. Biochem.* 43, 351–358. <https://doi.org/10.1016/j.soilbio.2010.10.023>
- Widenfalk, L.A., Leinaas, H.P., Bengtsson, J., Birkemoe, T., 2018. Age and level of self-organization affect the small-scale distribution of springtails (Collembola). *Ecosphere* 9, e02058. <https://doi.org/10.1002/ecs2.2058>
- Widenfalk, L.A., Malmström, A., Berg, M.P., Bengtsson, J., 2016. Small-scale Collembola community composition in a pine forest soil – Overdispersion in functional traits indicates the importance of species interactions. *Soil Biol. Biochem.* 103, 52–62. <https://doi.org/10.1016/j.soilbio.2016.08.006>
- Yan, S., Singh, A.N., Fu, S., Liao, C., Wang, S., Li, Y., Cui, Y., Hu, L., 2012. A soil fauna index for assessing soil quality. *Soil Biol. Biochem.* 47, 158–165. <https://doi.org/10.1016/j.soilbio.2011.11.014>
- Yang, C., Ji, Y., Wang, X., Yang, C., Yu, D.W., 2013. Testing three pipelines for 18S rDNA-based metabarcoding of soil faunal diversity. *Sci. China Life Sci.* 56, 73–81. <https://doi.org/10.1007/s11427-012-4423-7>
- Yang, Z., Rannala, B., 2016. Species Identification by Bayesian Fingerprinting: A Powerful Alternative to DNA Barcoding. *bioRxiv*. <https://doi.org/10.1101/041608>
- Yi, Z., Jinchao, F., Dayuan, X., Weiguo, S., Axmacher, J.C., 2012. A Comparison of Terrestrial Arthropod Sampling Methods. *J. Resour. Ecol.* 3, 174–182. <https://doi.org/10.5814/j.issn.1674-764x.2012.02.010>
- Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C., Ding, Z., 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol. Evol.* 3, 613–623. <https://doi.org/10.1111/j.2041-210X.2012.00198.x>

- Zaitsev, A.S., van Straalen, N.M., Berg, M.P., 2013. Landscape geological age explains large scale spatial trends in oribatid mite diversity. *Landscape Ecol.* 28, 285–296. <https://doi.org/10.1007/s10980-012-9834-0>
- Zhang, F., Deharveng, L., 2015. Systematic revision of Entomobryidae (Collembola) by integrating molecular and new morphological evidence. *Zool. Scr.* 44, 298–311. <https://doi.org/10.1111/zsc.12100>
- Zhang, F., Sun, D.-D., Yu, D.-Y., Wang, B.-X., 2015. Molecular phylogeny supports S-chaetae as a key character better than jumping organs and body scales in classification of Entomobryoidea (Collembola). *Sci. Rep.* 5. <https://doi.org/10.1038/srep12471>
- Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., Tang, M., Fu, R., Li, J., Huang, Q., 2013. Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience* 2, 4. <https://doi.org/10.1186/2047-217X-2-4>